Am J Hum Genet 27:348-364, 1975

Ethnic Variation in Genetic Disease: Possible Roles of Hitchhiking and Epistasis

D. K. WAGENER[1] AND L. L. CAVALLI-SFORZA[1]

INTRODUCTION

The incidence of genetic disease is receiving increased attention with the development of techniques of genetic prophylaxis because it affects the cost-benefit ratio of that procedure. This has emphasized interest in ethnic variation for genetic disease, a known phenomenon [1–6] whose origin has been the subject of investigation for specific diseases.

On the simplest assumption, the incidence of genetic disease is determined by a balance of mutation and selection. If so, variation of incidence between populations must take its origin from differences in mutation rates and/or differences in selection rates between populations. Variation may exist in selection coefficients against homozygotes and for/against heterozygotes. It is worth noting that even moderate differences in the selection coefficients of heterozygotes, often beyond the possibility of direct estimation, can cause relatively large differences of incidence [6].

But the assumption of equilibrium may be incorrect, especially if mutation or selection rates are changing or have recently changed. Also, drift is expected to cause at least some variation of gene frequency with a distribution which is predictable, given enough demographic information. Finally, interactions between genes may be important, causing gene frequencies to vary from mutation-selection equilibrium for one locus.

Explanations given for specific diseases have been based on either differences in selection rates between populations or drift. There is no major dispute as to the cause of the variation in incidence for sickle cell anemia, thalassemia, or G6PD which are primarily determined by local variation of selection coefficients. For Tay Sachs disease, different investigators have opted for founder effect [7–9] or local variation in selection for the heterozygote [10–12]. Founder effect is, of course, a special case of drift.

In a series of papers we will consider the explanations mentioned above. This paper considers gene interactions which would affect recessive diseases. We will look at the consequences of linkage or epistasis between the deleterious recessive and another gene which is being selected for in the population for other reasons.

Received July 15, 1974; revised November 20, 1974.

This work was supported by grant GM 20467-01 from the National Institutes of Health and grant AT(04-3)326 from the Atomic Energy Commission.

[1] Department of Genetics, Stanford University, Stanford, California 94305.

DATA

Tables 1 and 2 show the results of surveys involving phenylketonuria and cystic fibrosis. Studies prior to 1960 were generally excluded because tests for early detection either had not been developed or were not always reliable, therefore limiting the scope of those surveys. Surveys involving fewer than 20,000 births were avoided when larger ones were available for the same ethnic group. The techniques for collection of the remaining studies were heterogeneous; that is, most were voluntary, some involved surveys of death records, and others employed the use of case reports. Standard formulas were used in some surveys to

TABLE 1

INCIDENCE OF PHENYLKETONURIA FOR VARIOUS ETHNIC GROUPS

Region and Reference	No. Births	Incidence	Gene Frequency
Alabama:*			
Whites	392,370	1/17,000	.0077
Blacks	215,803	1/215,803	.0022
California [13]:			
Whites	2,038,945	1/17,134	.0076
Blacks	219,023	1/219,023	.0021
Georgia:†			
Whites	611,307	1/14,910	.0082
Blacks	312,762	1/156,381	.0025
Massachusetts:			
[14]	600,000	1/14,286	.0084
[15]	400,000	1/13,000	.0088
United States [16]	1,914,024	1/18,000	.0074
Australia:			
[17]	333,777	1/9,272	.0104
[18]	100,000	1/14,286	.0084
Austria [13]	247,743	1/8258	.0110
Belgium [13]	38,416	1/4,802	.0144
Canada:			
[19]	36,000	1/18,000	.0075
[20]	119,990	1/10,908	.0100
Ireland [21]	62,856	1/5,000	.0141
Israel [22]	178,174	1/25,000	.0063
Japan:			
[23]	5,419	1/58,825	.0048
[24]	...	1/45,000–1/65,000	.0047–.0040
New Zealand [25]	178,174	1/25,453	.0063
Scotland:			
[26]	80,000	1/4,700	.0146
[13]	350,000	1/6,604	.0123
Switzerland [13]	241,141	1/20,095	.0071
West Germany [27]	100,000	1/11,000	.0095

* Personal communication, Alabama Department of Public Health, 1973.

† E. Blake, personal communication, Georgia Department of Human Resources, 1973.

TABLE 2

INCIDENCE OF CYSTIC FIBROSIS FOR VARIOUS ETHNIC GROUPS

Region and Reference	No. Births	Incidence	Gene Frequency
Hawaii [28]:*			
Oriental	242,354	1/90,000	.0033
Caucasian	242,354	1/3,800	.0162
New England [29]	1,497,300	1/2,400	.0204
Ohio:			
Erie County [30]	121,981	1/2,400	.0204
[31]	742,163	1/3,750	.0163
Africa, Negro [32]	...	1/77,000	.0036
Australia [33]	362,732	1/2,250	.0211
England:			
[34]	188,613	1/2,400	.0204
[35]	546,764	1/4,100	.0156
Israel, Ashkenazi [36]	68,000	1/5,000	.0141
Sweden [37]	870,032	1/5,900	.0130

* This study includes 81,479 "pure" Caucasians and 96,968 "pure" Orientals; the rest are of mixed blood.

differentiate admixture between populations and compensate for consanguinity within populations.

A further note should be added to the phenylketonuria surveys, as reported by Hsia [38]. In the District of Columbia the mandatory mass screening procedure between 1967 and 1970 was discontinued due to the lack of cases. Over 84 of 100 births in these hospitals are nonwhite, yielding an apparent frequency of phenylketonuria of less than 1/100,000 births among nonwhites (gene frequency $< .0032$) and a rate of 1/13,000 among whites (comparable to other surveys).

The primary statistics computed from these data are the mean gene frequency and Wahlund's variance between populations, which were estimated as follows: $\bar{p} = \Sigma N_i p_i / \Sigma N_i$ and $f = \sigma^2/\bar{p}(1 - \bar{p}) = \Sigma N_i(p_i - \bar{p})^2/\bar{p}(1 - \bar{p})\,\Sigma N_i$, where N_i is the total size of population i, p_i is the gene frequency of population i, and the summations are taken over $i = 1, 2, \ldots, k$.

Both phenylketonuria and cystic fibrosis are relatively common among Caucasian populations and apparently rare among non-Caucasian populations. The surveys within major racial groups (Caucasian, Negroid, Oriental) were pooled to obtain the Wahlund's variance between the major racial groups, assuming relative sizes of .48, .14, and .38 [39]. The overall mean gene frequency and the f statistic are, respectively, 6.0×10^{-3} and 8.94×10^{-4} for phenylketonuria and 1.02×10^{-2} and 4.96×10^{-3} for cystic fibrosis. This assumes that the three major racial groups are homogeneous. This is not so true for phenylketonuria, the f statistic between European populations being 6.82×10^{-4}. For cystic fibrosis the f statistic between European populations (8.76×10^{-4}) is several times smaller than the f statistic between racial populations, therefore indicating less racial heterogeneity. It should be acknowledged that the number of populations

sampled is small and geographically biased. Moreover, data of variation within major racial groups are available only for Europeans.

Cystic fibrosis has a geographic distribution not unlike that of phenylketonuria, but the estimated f is somewhat larger. While this discrepancy may fall within limits of error due to sampling, new evidence for possible existence of other evolutionary factors has been suggested by Schanfield et al. [40]. An association between cystic fibrosis and a common Caucasian Gm haplotype, *fb*, has been noted. This association rests on two independent kinds of evidence: a disproportionate frequency of cystic fibrotic children with the *fb* allele from certain matings of carrier parents (i.e., one parent heterozygous and the other homozygous for *fb*) and also chemical interaction between the cystic fibrotic factor and IgG gammaglobulins [41]. The simplest explanations for this statistical association, apart from social or ethnic stratification, are linkage and epistasis. We will consider these two explanations in the next sections.

Other possible explanations have been considered, such as decreased penetrance, multiple loci, heterozygote advantage, and genetic drift. Rao and Morton [42] have recently concluded that drift may be a significant factor. We will discuss their approach and the general role of drift in recessive diseases in another paper [43]. Conneally et al. [5] have suggested that heterozygote advantage may be the cause of the increased gene frequency. This may be true especially of those populations that show higher incidences of the disease.

HITCHHIKING

The simplest explanation for the association reported between cystic fibrosis and Gm loci would be linkage. If a new selectively advantageous mutant is introduced into a population, the alleles of genes linked to the mutated locus will undergo an increase in frequency, at least temporarily, as the new mutant becomes established in the population. This increase in frequency is the result of the "hitchhiking" relationship of these certain alleles to the new mutant. This model has been studied [44–46], both deterministically and stochastically, to determine the decrease in the probability of the hitchhiking alleles being lost. The problem we wish to address is the magnitude of the apparent, though temporary, increase in gene frequency of a hitchhiking recessive lethal allele. We will not be concerned with the fixation probability since the fixation of a recessive lethal is impossible. Further, since the selection for a recessive lethal is in fact strong, we cannot use the approximations suggested in the above papers.

Consider a chromosome with two loci, A and B. Gametes undergo mutation, unite at random, undergo selection, and finally recombine to form new gametes. Let A be a locus with two alleles, A and a, where A has an additive (s) or multiplicative (w) selective advantage. At the B locus two alleles, B and b, are present, where b is the recessive lethal allele. Further, μ is the mutation rate at the B locus from B to b, and r is the recombination frequency between the two loci. The gametes are AB, Ab, aB, and ab with frequencies x_1, x_2, x_3, and x_4. A series of

recurrence equations can be set up to describe the gamete frequencies of the above model (see Appendix).

Of particular interest to us in this model is the frequency of the deleterious allele b. If we denote this frequency as $q = x_2 + x_4$, then the change in q between successive generations may be written in terms of the gamete frequencies of the parental generation. For the additive model,

$$\Delta_a q = \frac{1}{T_{add}} \{(1 - 2q)(1 - \mu)[pq + s(x_1x_2 - x_3x_4)] - (q - \mu)[p^2 + sp(x_1 - x_3)]\}, \quad (1)$$

where $p = 1 - q$ is the frequency of B and T_{add} is the normalization factor due to the additive selection on the parental generation. And, similarly, for the multiplicative model,

$$\Delta_m q = \frac{1}{T_{mult}} \left\{(1 - 2q)\left(\frac{1 - \mu}{2}\right)[w^2x_3x_4 + w(x_1x_4 + x_2x_3) + x_1x_2] - (q - \mu)[x_1 + wx_3]^2\right\}, \quad (2)$$

where T_{mult} is the normalization factor for the multiplicative model.

Note that the first terms (in brackets) of equations (1) and (2) include zygotes homozygous for B, and the second terms include zygotes heterozygous at the B locus. The coefficients of the second terms are always positive since $q \geqq \sqrt{\mu} > \mu$. The coefficients of the first terms are positive for $q < .5$. Because the allele b is lethal in the homozygous state, without loss of generality we may consider q always less than .5, and, therefore, the first terms are always positive. The change in frequency of the lethal recessive depends on the difference between the two terms. As long as the A allele is primarily associated with the b allele, Δq will be positive (i.e., the first term is large). When the B allele becomes associated with the A allele, the second term becomes large and Δq becomes negative. However, it is of interest to note that the recombination rate does not enter into this equation. This will be verified later by a computer model.

Because the above models are not amenable to full algebraic analysis, computer models were made to study the dynamics of these systems. The initial population for the models is made up of N individuals. At the B locus the population is at mutation-selection equilibrium. At the A locus the a allele is the only allele present except for one Ab gamete. The initial frequency vector is then $0,\ 1/2N,\ 1 - \sqrt{\mu},\ \sqrt{\mu} - (1/2N)$. The initial introduction of A into the population may be the unique mutation of the new allele A in a b chromosome or perhaps the introduction of such a chromosome through migration.

With $\mu = 10^{-5}$ and $N = 10^5$ and 10^3, various combinations of s (or w) and r were tried. Figure 1 shows sample runs for the additive model. Table 3 gives

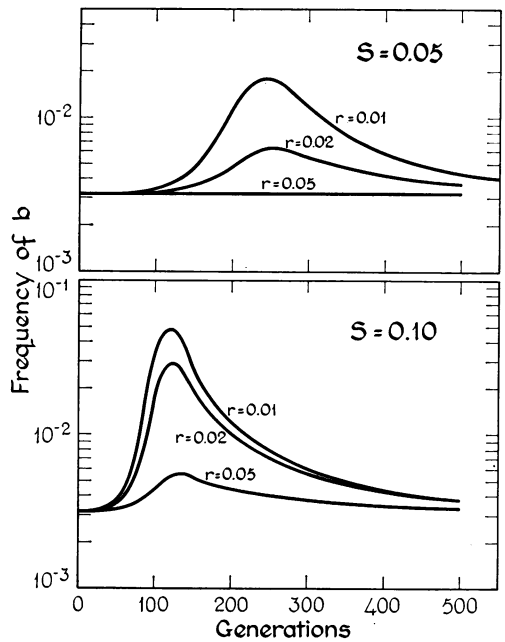

FIG. 1.—Frequency of lethal recessive allele over time from hitchhiking model with initial population size of 100,000 and mutation rate of 10^{-5}. b = recessive lethal allele at B locus; s = selection at A locus; r = recombination rate between A and B loci.

various peak values for the frequency of the deleterious recessive allele under both the additive and multiplicative models. The frequency q increases in both models to a peak due to its hitchhiking relationship with the A allele.

Figure 1 indicates that the magnitude of selection (s) on A and the number of generations required to reach a peak value are inversely related. However, the

TABLE 3

MAXIMUM FREQUENCIES FOR HITCHHIKING RECESSIVE ALLELE

SELECTION PARAMETER	RECOMBINATION FREQUENCY .05	.02	.01	.005	EFFECTIVE POPULATION SIZE	TIME (GENERATIONS)
Additive (s):						
.10	0.0053	0.0253	0.0465	0.0592	100,000	125
	0.0175	0.0442	0.0591	0.0694	1,000	75
.05	0.0032	0.0062	0.0174	0.0270	100,000	250
	0.0047	0.0153	0.0258	0.0333	1,000	150
.02	...	0.0032	0.0034	0.0061	100,000	600
	...	0.0040	0.0064	0.0107	1,000	350
Multiplicative (w):						
.909	0.0045	0.0427	0.0454	0.0566	100,000	130
.952	...	0.0057	0.0162	0.0626	100,000	260

NOTE.—Equilibrium value = 0.0032; mutation rate = 10^{-5}.

recombination frequency does not affect the time of the peak. Instead, recombination frequency and the maximum value of the peak are inversely related. The change in gene frequency between generations [see equations (1) and (2)] in two-loci models does not depend on r. However, the change in frequency of gametes with time depends on r, which determines the rate at which the B allele becomes associated with the A allele, thereby increasing the second terms of equations (1) and (2). The initial value for x_2 is determined by N: when N is small, x_2 is large, the base upon which the cumulative hitchhiking effect is built is greater, and the peak value is larger.

From the computer models, the disequilibrium ($D = x_2x_3 - x_1x_4$) between the two loci can be calculated at each iteration step. Figure 2 shows the change of D with time. Note that throughout this process the gametes are at disequilibrium. Therefore, we cannot write the gametic frequencies as a simple product of the allelic frequencies as Ewens [47] did with his analysis of the evolution of dominance.

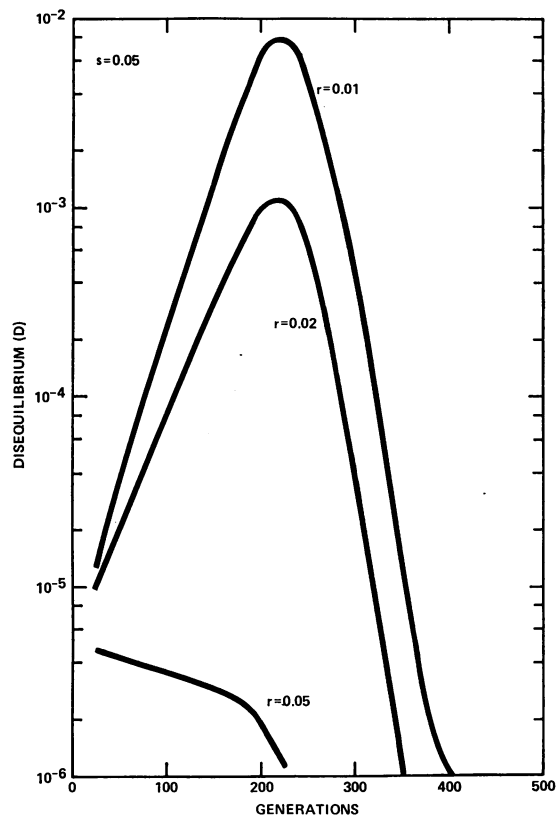

FIG. 2.—Disequilibrium (D) over time from hitchhiking model (additive selection) with initial population size of 100,000 and mutation rate of 10^{-5}. s = selection at A locus; r = recombination rate between A and B loci; $D = x_2x_3 - x_1x_4$.

Having seen that the dynamics of the lethal recessive could be profoundly affected, we now turn to the evolution of the selectively advantaged A allele and compare it with the evolution of a similar allele in a one-locus model. For a one-locus model with selective fitnesses of $1 + s$, 1, and $1 - s$ for the genotypes AA, Aa, and aa, respectively, standard results from the continuous approximation give us the following kinetic equation [6]:

$$P = \frac{P_0 e^{sT}}{Q_0 + P_0 e^{sT}}, \tag{3}$$

where T is the number of generations from the beginning, P is the frequency of the new allele A at time T, and P_0 and Q_0 are the initial frequencies of A and a. Using this equation we compared the predicted times, in generations, for the frequency of A to reach selected values and the times at which these frequencies were attained in the model. As expected, for s and w values not too small, the dynamics of the A locus were not influenced in an important way by its linkage with a rare recessive deleterious allele. Further, it was noted empirically from the computer runs that the time at which the peak occurs for q coincides with the time at which the frequency of A is $50\% \pm 5\%$. This allows us, therefore, to predict the time of maximum hitchhiking effects on b, given the additive selective advantage of A.

From figure 1 one notes that the greater the recombination between the deleterious locus and the transient polymorphic locus, the smaller the hitchhiking effect. The conditions on r for increase of the lethal recessive can be found in a model by Bodmer and Parsons [48] (see also [49]). They derived conditions for the mutual increase of rare alleles at each of two loci. The condition given by their symmetric fitness model is

$$r < \frac{s}{1 + s}, \tag{4}$$

where s is the overdominance of fitness for the balanced heterozygote. In the hitchhiking model, there are only three gametes present after the introduction of the A allele (Ab, aB, and ab). The possible genotypes are shown in table 4 with their respective fitnesses. Therefore, in the initial population the overdomi-

TABLE 4

GENOTYPES AND RESPECTIVE FITNESSES IN HITCHHIKING MODEL

	Ab	*aB*	*ab*
Ab	0	$1 + s$	0
aB	$1 + s$	1	1
ab	0	1	0

nance of the balanced heterozygote ($AaBb$) is equivalent to the additive advantage s of the hitchhiking model. The inequality (4) may be used to calculate conditions of mutual increase of the rare alleles A and b in the initial population of the hitchhiking model. This hitchhiking model mimics the symmetric overdominance model at least to some extent [48], although differences between them should be noted. In the symmetric overdominance model, all four gametes may be present when new equilibria are reached. However, the selection scheme of the hitchhiking model is neither balanced nor is the marginal selection of either locus heterotic. The hitchhiking model, therefore, does not have any equilibria where all four gametes are present. But temporary peak values for the frequency of rare alleles are attained.

Of course, as r approaches zero the situation changes. And for $r = 0$, the problem reduces to a three-allele situation (Ab, aB, and ab). This system would eventually reach a new mutation-selection equilibrium with a gene frequency of q different from the initial population.

UNLINKED EPISTATIC MODEL

Another model which produces a marked increase in the recessive lethal gene frequency is that of an epistatic interaction with an unlinked allele which spreads through the population [50]. If the loci are unlinked, the gametes undergo mutation and then selection. The recurrence equations for a multiplicative epistatic effect are given in the Appendix.

In the hitchhiking model the beginning and ending equilibrium lethal gene frequencies were the same. However, in the epistatic model the equilibria are different because the marginal selection is different. When the frequency of A is zero, the marginal selection of the B locus for genotypes BB, Bb, and bb is $1 - s$, $1 - s$, and 0, respectively, so the mutation-selection equilibrium frequency of b is $\sqrt{\mu}$. When the frequency of A is one, the marginal selection at the B locus for genotypes BB, Bb, and bb is $1 + s$, $(1 + s)(1 + e)$, and 0, where e is the epistatic parameter. The equilibrium lethal gene frequency for this balanced selection model is $e/(1 + 2e)$. We will choose values of e such that q increases. In these cases the frequency of the lethal allele will change monotonically to its new equilibrium value. If e is chosen so that the final equilibrium is less than or equal to the initial q_0, the increase in q will be temporary, as in the hitchhiking model.

As before, a computer model was used with the same initial gene frequencies as in the hitchhiking model. Figure 3 displays various results from the model. As expected, the final equilibrium frequencies depend upon the parametric values of e, not s (selection). However, s determines the time course of the process. If the frequency of A is studied separately, it is found that the evolution of this locus, as before, is essentially a one-locus model. Therefore the effects due to the rare allele b are minimal.

Figures 4 and 5 show the frequency of b (q) versus the frequency of A (P). From figure 4 we note that as selection is held constant and the epistatic parameter is allowed to vary, an increase in e will increase the frequency of q for any

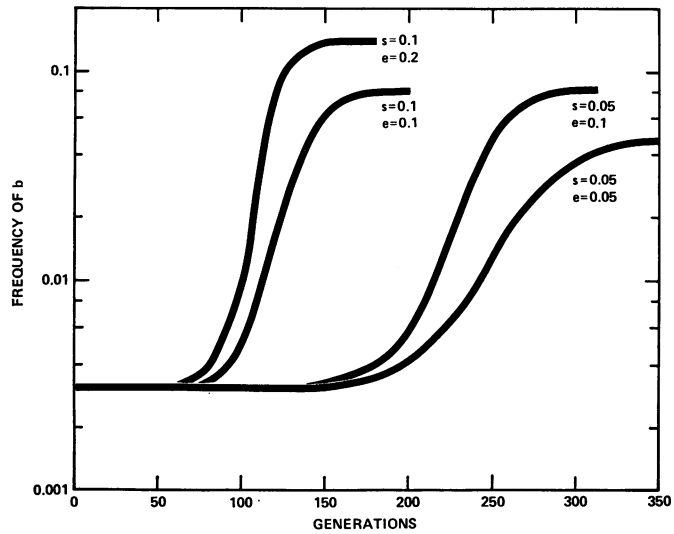

FIG. 3.—Frequency of lethal recessive allele *b* over time from unlinked epistatic model with initial population size of 100,000 and mutation rate of 10^{-5}. s = selection at A locus; e = epistatic parameter between A and B loci.

given value of *P*. Figure 5 shows that if the epistatic interactions are held constant, as selection decreases the frequency q increases for any given frequency of *A*.

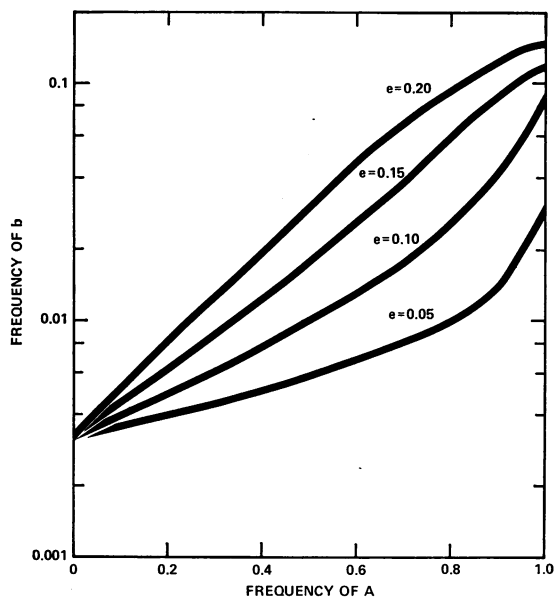

FIG. 4.—Frequency of lethal recessive allele *b* versus frequency of advantageous allele *A* from unlinked epistatic model with selective advantage of *A* held constant at .1, initial population size of 100,000, and mutation rate of 10^{-5}.

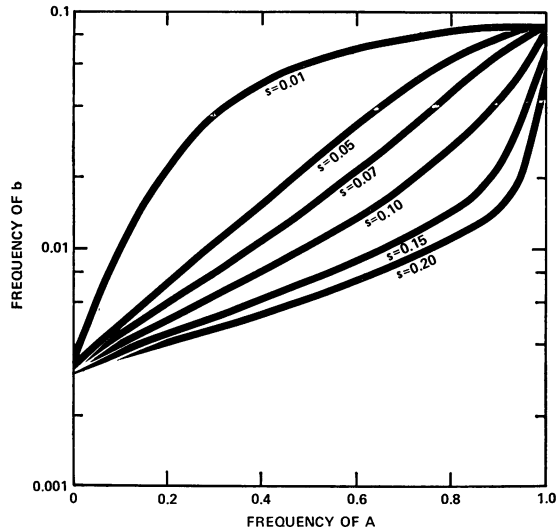

FIG. 5.—Frequency of lethal recessive allele *b* versus frequency of advantageous allele *A* from unlinked epistatic model with epistatic parameter held constant at .1, initial population size of 100,000, and mutation rate of 10^{-5}.

DISCUSSION OF MODELS

In both models presented, the association of the lethal recessive allele with a locus undergoing a transient polymorphism has greatly altered the dynamics of the lethal gene. The association causes the lethal recessive allele to mimic a heterotic situation, that is, to produce a directional increase in frequency to a level significantly greater than the mutation rate. However, the increase in frequency may be temporary, unlike a heterotic model.

The hitchhiking effect described here may be important for several cases in which new alleles may have strong advantageous effects. For instance, the multiplicity of haplotypes at both Gm and HL-A loci may indicate at least temporary strong favorable selection rates for new alleles. Lactose tolerance may have undergone favorable selection with an increased advantage as large as .04 among Caucasians [51]. In each of these cases, linkage with a rare lethal recessive allele would probably have little effect on the dynamics of the advantageous allele. However, the dynamics of the recessive allele may be profoundly affected.

In particular, let us examine the possible case of a hitchhiking association between the Gm and cystic fibrosis loci. The frequency of the cystic fibrosis gene is five to six times more frequent in Caucasian populations than in other racial groups. The indication that the cystic fibrotic allele may be nonrandomly associated with a particular Caucasian Gm haplotype, *fb,* suggests a possible hitchhiking relationship. In the Gm system we find the haplotypes which are racially specific have frequencies as high as .5 or greater (e.g., .76 for the *fb* haplotype in Caucasian populations [52]). Given the length of time over which the haplotype

has evolved from initial mutation of a to A, we can use equation (3) to predict the selective values required to increase the haplotype frequency to 50%. In table 5 we have chosen several intervals of time which may be of particular significance in racial evolution: 30,000 years ago is estimated to be a lower limit for the separation of the major ethnic groups, and 10,000 years ago is approximately the beginning of the neolithic expansion out of the Middle East. Equation (4) is then used to estimate the maximum recombination distance within which hitchhiking genes would be expected to have their gene frequencies temporarily altered. These estimates indicate that the two loci must be closely linked. However, if one is willing to assume that the new haplotype was generated in the Caucasian population as recently as, say, 100 generations ago, a much looser linkage is tolerated because the selective advantage is larger. This demonstrates what was given earlier, namely, that for recombination r less than selection s, the frequencies of linked alleles around the locus of the spreading allele will undergo a temporary rise in the gene frequency.

Although the rise in frequency of the lethal allele is temporary, the length of time that the gene frequency is increased may be quite long. For instance, consider the time during which the gene frequency is at least double the mutation-selection equilibrium frequency. For $s = .05$ and $r = .01$, the period is about 225 generations; for $s = .10$ and $r = .01$, 250 generations; and for $r = .02$, 200 generations.

Family data on the linkage between the Gm and cystic fibrosis loci are limited (M. S. Schanfield, unpublished observation) and insufficient to prove linkage. Given the very high heterogeneity of Gm frequencies in populations, selection coefficients on the order of .10 may be involved for these alleles. This is then the maximum recombination value expected to give some hitchhiking effect if Gm is linked to cystic fibrosis. The linkage thus required may be too high even with the very limited linkage data available.

Using the epistatic model, one can estimate sets of values for epistasis and selection which will produce the frequencies of *fb* and cystic fibrosis that have been

TABLE 5

MAXIMUM RECOMBINATION DISTANCES FOR HITCHHIKING EFFECTS

Initial Event to *Ab* (No. Years Ago)	Time (Generations)	Selective Advantage s	Relative Increase ($\times q_E$)	Recombination Distance r
30,000	1,000	.01	2	0.004
10,000	350	.03	5	0.004
			2	0.023
3,000	100	.10	5	0.072
			2	0.093

NOTE.—Estimates are given for various anthropological dates and current frequencies of Gm haplotypes. The parameter for additive selection (s) is then chosen so that the peak frequency of the lethal recessive is at least two to five times the initial equilibrium frequency.

observed, namely, .776 and .015–.020 (table 6). The association of IgG and the cystic fibrotic factor may provide insights into problems of epistasis. If linkage is not the explanation of the Gm–cystic fibrosis association, however, ethnic stratification may have to be ruled out before an epistatic explanation is accepted.

The frequency with which heterogeneity due to hitchhiking will be found depends on the density of deleterious mutations. We can get an estimate of the probability of hitchhiking affecting a lethal gene by using the estimates of 1.1 lethal equivalents per haploid human genome [6] and 50 chiasma per genome [53]. Since each chiasma corresponds to a haploid distance of $r = .5 = 50$ cM, we would expect $1.1/(50 \times 50) = .0004$ lethals per centimorgan. Consider a polymorphic locus with an additive selective advantage of $s = .04$. Then for genes within a recombinational distance of $r = \pm .04$, we would expect hitchhiking effects. And, in particular, the probability that a lethal allele is present among these genes is then $.0004 \times 8 = .0032$.

SUMMARY

The high incidence of some genetic diseases in certain ethnic groups is important in planning of medical genetic programs. Simple interaction models predict that at least some lethal recessive alleles will have "hitchhiked" to increased frequencies because of linkage to genes whose alleles have been favored by selection for other reasons in certain populations. In the absence of linkage or epistasis with a gene favored by selection, heterozygote advantage for a recessive lethal may produce the same phenomenon. In the hitchhiking model (linkage), the increase in the gene frequency is temporary, but the length of time that the increased gene frequency is at least double the base frequency may be quite long. Changes in gene frequency for the unlinked epistatic model result in a new equilibrium with a possibly higher gene frequency.

The most likely chromosomal regions in which hitchhiked lethal recessives would be found are in the vicinity of genes whose allelic frequencies vary substantially

TABLE 6

VALUES FOR EPISTASIS AND SELECTION PRODUCING OBSERVED FREQUENCIES OF *fb* AND CYSTIC FIBROSIS

Generations	Selective Advantage s	Epistatic Influence e	Lethal Recessive Frequency
85	.15	.10	.018
125	.10	.10	.023
130	.10	.07	.016
265	.05	.05	.020
265	.05	.04	.015
450	.03	.03	.017

NOTE.—Observed frequencies of *fb* and cystic fibrosis in Caucasian populations are .776 and .015–.020, respectively.

among human racial groups (e.g., Gm, Rh, Duffy, lactose tolerance, or HL-A). There will be a hitchhiking effect if recombination distance is less than the selective advantage.

The closer the linkage of two loci, the easier hitchhiking effects will be to detect. Hitchhiking is suggested by nonrandom association of the recessive disease and one of the selected markers, as in the case of Gm and cystic fibrosis. However, there is so far insufficient evidence of linkage between them. More pedigree information is necessary than is now available.

ACKNOWLEDGMENTS

We thank Dr. M. Schanfield for many useful and informative discussions. We also thank Mr. S. Daiger for his useful suggestions.

APPENDIX

Given four types of gametes (AB, Ab, aB, and ab) with frequencies x_1, x_2, x_3, and x_4 in the parental generation and x_1', x_2', x_3', and x_4' in the offspring generation, the transition from the parental to offspring frequencies for the additive hitchhiking model is then given by:

1. *Mutation.* The intermediate gamete frequencies $\tilde{x}$ become $x_1(1-\mu)$, $x_2+\mu x_1$, $x_3(1-\mu)$, and $x_4+\mu x_3$. Then mating occurs at random.

2. *Selection.* If the selection on the diploid individuals is additive, then the fitness matrix may be written as

$$W_{add}=\begin{pmatrix} 1+s & 1+s & 1 & 1 \\ 1+s & 0 & 1 & 0 \\ 1 & 1 & 1-s & 1-s \\ 1 & 0 & 1-s & 0 \end{pmatrix},$$

and the resulting intermediate gamete frequencies $\tilde{\tilde{x}}$ become $\tilde{x}W_{add}\tilde{x}'$. At this stage the terms of order μ^2 are ignored.

3. *Recombination.* Recombination occurs in the double heterozygote. The resulting gamete frequencies after 1 generation are

$$x_1'=\frac{1}{T_a}\left[x_1(1+s(1-\mu)(x_1+x_2)-\mu)+r(1-\mu)D\right]$$

$$x_2'=\frac{1}{T_a}\left[x_2(x_1+x_3)+sx_1x_2+\mu(x_1+x_3)(x_1-x_2)\right.$$
$$\left.+s\mu x_1(x_1-x_2)-r(1-\mu)D\right]$$

$$x_3'=\frac{1}{T_a}\left[x_3(1-s(1-\mu)(x_1+x_2)-\mu)-r(1-\mu)D\right]$$

$$x_4'=\frac{1}{T_a}\left[x_4(x_1+x_3)-sx_3x_4+\mu(x_1+x_3)(x_3-x_4)\right.$$
$$\left.-s\mu x_3(x_3-x_4)+r(1-\mu)D\right],$$

where T_a is the normalization factor so that $\Sigma x_i' = 1 = \Sigma x_i$, and D is the disequilibrium written in terms of the parental generation $x_2x_3 - x_1x_4$. Note that $x_1 + x_2$ is the gene frequency of A; $x_1 + x_3$ is the frequency of B; and r is the recombination distance.

The multiplicative model is similar, but the fitness matrix is

$$W_{mult} = \begin{pmatrix} 1 & 1 & w & w \\ 1 & 0 & w & 0 \\ w & w & w^2 & w^2 \\ w & 0 & w^2 & 0 \end{pmatrix}.$$

For the epistatic model the iteration scheme was the same as in the hitchhiking model, with $r = .5$ and selection matrix

$$W_{epis} = \begin{pmatrix} 1+s & (1+s)(1+e) & 1 & 1+e \\ (1+s)(1+e) & 0 & 1+e & 0 \\ 1 & 1+e & 1 & 1 \\ 1+e & 0 & 1 & 0 \end{pmatrix}.$$

REFERENCES

1. Damon A: Race, ethnic groups and disease. *Soc Biol* 16:69–80, 1969
2. McKusick VA: *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes*, 3d ed. Baltimore, Johns Hopkins Press, 1971
3. Lobeck CC: Cystic fibrosis, in *The Metabolic Basis of Inherited Disease*, edited by Stanbury JB, Wyngaarden JB, Fredrickson DS, New York, McGraw-Hill, 1972, pp 1605–1623
4. McKusick VA: Ethnic distribution of disease in non-Jews. *Isr J Med Sci* 9:1375–1382, 1973
5. Conneally PM, Merritt AD, Yu P: Cystic fibrosis: population genetics. *Tex Rep Biol Med* 31:639–650, 1973
6. Cavalli-Sforza LL, Bodmer WB: *Genetics of Human Populations*. San Francisco, Freeman, 1971
7. Chase GA, McKusick VA: Founder effect in Tay Sachs disease. *Am J Hum Genet* 24:339–340, 1972
8. McKusick VA: Clinical genetics at a population level: the ethnicity of disease in the United States. *Ala J Med Sci* 3:408–424, 1966
9. McKusick VA: The ethnic distribution of disease in the United States. *J Chronic Dis* 20:115–118, 1967
10. Myrianthopoulos NC, Naylor AF, Aronson SM: Founder effect in Tay Sachs disease unlikely. *Am J Hum Genet* 24:341–342, 1972
11. Myrianthopoulos NC, Aronson SM: Population dynamics of Tay Sachs disease: I. Reproductive fitness and selection. *Am J Hum Genet* 18:313–327, 1966
12. Myrianthopoulos NC, Aronson SM: Population dynamics of Tay Sachs disease. II. What confers the selective advantage upon the Jewish heterozygote?, in *Proceedings 4th International Symposium on Sphingolipidoses*, New York, Plenum, 1972, pp 561–570
13. Cunningham GC: Phenylketonuria testing: its role in pediatrics and public health. *CRC Crit Rev Clin Lat Sci* 2:45–101, 1971
14. Levy HL, Karolkewicz V, Houghton SA, MacCready RA: Screening the "normal" population of Massachusetts for PKU. *N Engl J Med* 282:1455–1458, 1970

15. EFRON ML: Disorders of the ornithine-urea cycle, in *Amino Acid Metabolism and Genetic Variation*, edited by NYHAN NL, New York, McGraw-Hill, 1967, pp 219–224
16. BERMAN JL, CUNNINGHAM GC, DAY RW, FORD R, HSIA DYY: Causes for high phenylalanine with normal tyrosine in newborn screening programs. *Am J Dis Child* 117:54–65, 1969
17. PITT D: Preliminary results of screening for metabolic errors in Australia. *Med J Aust* 1:1073, 1970
18. TURNER B, BROWN DA: Amino acid excretion in infancy and early childhood. *Med J Aust* 1:11–14, 1970
19. CLOW C, SCRIVER CR, DAVIES E: Results of mass screening for hyperaminoacidemias in the newborn infant. *Am J Dis Child* 117:48–53, 1969
20. HANLEY WB, PARTINGTON MW, RATHBUN JC, AMIES CR, WEBB JF, MOORE JE: The newborn phenylketonuria screening program in Ontario. *Can Med Assoc J* 101: 185–190, 1969
21. CAHALANE SF: PKU: mass screening of newborns in Ireland. *Arch Dis Child* 43: 141–144, 1968
22. SZEINBERG A, COHEN BE, GOLAN R, PELED I, LAVI U, CRISPIN M: Persistent mild hyperphenylalaninemia in various ethnic groups in Israel. *Am J Dis Child* 118:559–564, 1969
23. TANAKA K, MATSUNAGA E, HANDA Y, MURATA T, TAKEHARA K: Phenylketonuria in Japan. *Jap J Hum Genet* 6:65–77, 1961
24. HASHIMOTO S: Prevention of birth defects. *Popular Med* 4, suppl.:196–202, 1973
25. BECROFT DMO, HORN CR: Guthrie screening test for phenylketonuria: a report of two years participation in the national programme. *NZ Med J* 69:212–215, 1969
26. MCBEAN MS: Treatment of classical phenylketonuria. *Arch Dis Child* 43:1–7, 1969
27. BICKEL H: Treatment of phenylketonuria, in *Conference on Phenylketonuria and Allied Metabolic Diseases*, edited by ANDERSON JA, SWAIMAN MK, Washington, D.C., Children's Bureau, 1967, pp 99–115
28. WRIGHT SW, MORTON NE: Genetic studies on cystic fibrosis in Hawaii. *Am J Hum Genet* 20:157–169, 1968
29. KRAMM ER, CRANE MM, SIRKEN MG, BROWN ML: A cystic fibrosis pilot survey in three New England states. *Am J Public Health* 52:2041–2057, 1962
30. SULTZ HA, SCHLESINGER ER, MOSHER WE: The Erie County survey of long-term childhood illness. II. Incidence and prevalence. *Am J Public Health* 58:491–498, 1968
31. STEINBERG AG, BROWN DC: On the incidence of cystic fibrosis of the pancreas. *Am J Hum Genet* 12:416–424, 1960
32. WRIGHT SW: Racial variation in the incidence of cystic fibrosis, in *Proceedings 5th International Conference on Cystic Fibrosis of the Pancreas*, Cambridge, Churchill College, 1969, pp 91–102
33. DANKS DM, ALLAN J, ANDERSON CM: A genetic study of fibrocystic disease of the pancreas. *Ann Hum Genet* 28:323–356, 1965
34. HALL BD, SIMPKISS MJ: Incidence of fibrocystic disease in Wessex, in *Proceedings 5th International Conference on Cystic Fibrosis of the Pancreas*, Cambridge, Churchill College, 1969, pp 82–90
35. PUGH RJ, PICKUP JD: Cystic fibrosis in the Leeds region: incidence and life expectancy. *Arch Dis Child* 42:544–545, 1967
36. LEVIN S: Fibrocystic disease of the pancreas, in *The Genetics of Migrant and Isolate Populations*, edited by GOLDSCHMIDT E, New York, Williams & Wilkins, 1963, pp 294–295
37. SELANDER P: The frequency of cystic fibrosis of the pancreas in Sweden. *Acta Paediatr* 51:65–67, 1962

38. HSIA DY: A critical evaluation of PKU screening. *Hosp Pract* 6(4):101–112, 1971
39. UNESCO: *Statistical Yearbook*. Belgium, UNESCO, 1966
40. SCHANFIELD MS, MCCOMBS ML, FUDENBERG HH: The distribution of immunoglobin Gm phenotypes among cystic fibrosis families (abstr.). *Am J Hum Genet* 25:65A, 1973
41. BOWMAN BH, KUROSKY A, FULLER GM, BARNETT DR, LANKFORD BJ, CARSON S: Cystic fibrosis: association of a small polypeptide with serum IgG (abstr.). *Am J Hum Genet* 26:17A, 1974
42. RAO DC, MORTON NE: Large deviations in the distribution of rare genes. *Am J Hum Genet* 25:594–597, 1973
43. WAGENER DK, CAVALLI-SFORZA LL: Ethnic variation of diseases: models of drift. In preparation
44. KOJIMA K, SCHAFFER HE: Survival process of linked mutant genes. *Evolution* 21:518–531, 1967
45. SMITH JM, HAIGH J: The hitch-hiking effect of a favourable gene. *Genet Res* 23:23–35, 1974
46. HILL WG, ROBERTSON A: The effect of linkage on limits to artificial selection. *Genet Res* 8:269–294, 1966
47. EWENS WJ: Further notes on the evolution of dominance. *Heredity* 20:443–450, 1965
48. BODMER WF, PARSONS P: Linkage and recombination in evolution. *Adv Genet* 11:1–100, 1962
49. BODMER WF, FELSENSTEIN J: Linkage and selection: theoretical analysis of the deterministic two-locus random mating model. *Genetics* 57:237–265, 1967
50. KOJIMA K, SCHAFFER HE: Accumulation of epistatic gene complexes. *Evolution* 18:127–129, 1964
51. CAVALLI-SFORZA LL: Origin and differentiation of human races, Huxley Memorial Lecture, in *Proceedings of the Royal Anthropological Institute of Great Britain and Ireland,* 1972, pp 15–25
52. SCHANFIELD MS, FUDENBERG HH: The anthropological usefulness of the IgA allotypic markers. In preparation
53. HULTEN M, LINDSTEN J: Cytogenetic aspects of human male meiosis. *Adv Hum Genet* 4:327–388, 1973