Am J Hum Genet 29:195–207, 1977

Selection-Mutation Balance for Two Nonallelic Recessives Producing an Inferior Double Homozygote

FREDDY BUGGE CHRISTIANSEN[1] AND OVE FRYDENBERG[2]

Many examples of hereditary human frailties which do not segregate in accordance with simple monogenic expectations are known. The abnormal variants are often rarer than expected, and their segregation appears to vary from family to family. Limited penetrance has been held responsible for this deficiency, implying an interaction between the supposed monogenic genotype and environmental factors. However, odd and irregular segregations are also expected when a trait depends on two or more genes. The latter explanation, though currently popular, reduces our knowledge of the population dynamics of a number of genetic disorders from a state of well explored theory to one of unwarranted guesses at a time when population genetics is most crucial for evaluation of the hazards of increasing mutation rates [1].

Inspired by the current debate about ionizing radiation and human illness, we have studied a model for a disorder caused by the simultaneous homozygosity of two recessive genes. This model is similar to those considered by Fisher [2] and Nei and Roychoudhury [3] in their discussions of evolution at duplicate loci. Although our model is no more simplified than those commonly accepted for single genes, it is still very crude. However, some surprising properties relevant to human genetics in general and to the question of mutational burdens in particular have been revealed.

THE MODEL

Definition and Basic Assumptions

Consider in a random mating population two loci, each with two alleles (A, a and B, b) and recombination frequency r. Assume that mutations are effectively unidirectional with frequencies μ for $A \rightarrow a$ and ν for $B \rightarrow b$. Assume further that all genotypes have normal phenotypes (with a fitness of unity), except the disabled double homozygote, $aabb$, whose fitness is $1 - s$ $(0 < s \leq 1)$. Define the gametic frequencies in a given generation as $x_1 = \text{freq}(AB)$, $x_2 = \text{freq}(Ab)$, $x_3 = \text{freq}(aB)$, and $x_4 = \text{freq}(ab)$. Then the gene frequencies are $p_A = \text{freq}(A) = x_1 + x_2$; $q_a = \text{freq}(a) = x_3 + x_4 = 1 - p_A$; $p_B = \text{freq}(B) = x_1 + x_3$; and $q_b = \text{freq}(b) =$

Received June 6, 1976; revised November 5, 1976.

This work was supported in part by grant NIH 10452-11 from the U.S. Public Health Service and by grant GB 37835 from the National Science Foundation.

[1] Department of Ecology and Genetics, University of Aarhus, Ny Munkegade, DK 8000, Aarhus, Denmark.

[2] Deceased April 7, 1975.

© 1977 by the American Society of Human Genetics. All rights reserved.

$x_2 + x_4 = 1 - p_B$. Possible departures from random association of alleles in the gametes will be measured by the linkage disequilibrium, $d = x_1x_4 - x_2x_3$, or by the squared gametic correlation, $\zeta^2 = d^2/(p_A q_a p_B q_b)$. The mean fitness of the population is $w = 1 - s x_4^2$.

Recurrence Equations

Assume that the frequencies of gametes in an arbitrary generation are x_1, x_2, x_3, and x_4. Neglecting mutation for the moment, this generation will produce gametes with the following frequencies:

$$x_1^* = (x_1 - rd)/w; \tag{1a}$$

$$x_2^* = (x_2 + rd)/w; \tag{1b}$$

$$x_3^* = (x_3 + rd)/w; \text{ and} \tag{1c}$$

$$x_4^* = (x_4 - sx_4^2 - rd)/w. \tag{1d}$$

However, mutation has changed these frequencies into the actual frequencies of the gametes of the next generation:

$$x_1' = (1 - \mu)(1 - \nu) x_1^*; \tag{2a}$$

$$x_2' = (1 - \mu) x_2^* + (1 - \mu) \nu x_1^*; \tag{2b}$$

$$x_3' = (1 - \nu) x_3^* + (1 - \nu) \mu x_1^*; \text{ and} \tag{2c}$$

$$x_4' = x_4^* + \mu\nu x_1^* + \mu x_2^* + \nu x_3^*. \tag{2d}$$

The equations (2a)–(2d) with x_1^*, x_2^*, x_3^*, and x_4^* from equations (1a)–(1d) constitute a system of recurrence equations linking the gametic frequencies in the offspring generation to those in the parent generation. By adding equations (2a) and (2b), we obtain a recurrence equation in the gene frequency of A from equations (1a) and (1b):

$$p_A' = (1 - \mu)(x_1^* + x_2^*) = (1 - \mu) p_A/w. \tag{3}$$

Similarily, we can get a recurrence equation in the frequency of B, so that the gene frequencies in the offspring generation are:

$$p_A' = (1 - \mu) p_A/w; \tag{4a}$$

$$p_B' = (1 - \nu) p_B/w. \tag{4b}$$

Stable Equilibrium when $\mu \neq \nu$

Neglecting the trivial equilibrium $p_A = p_B = 0$ and taking the ratio of equations (4a) and (4b) gives

$$\frac{p_A'}{p_B'} = \frac{1 - \mu}{1 - \nu} \frac{p_A}{p_B}, \tag{5}$$

which immediately generalizes to

$$\frac{p_A^{(n)}}{p_B^{(n)}} = \left(\frac{1 - \mu}{1 - \nu}\right)^n \frac{p_A}{p_B} \tag{6}$$

(i.e., the ratio of the gene frequencies of the alleles A and B evolves independently of selection). In the realistic situation where $\mu \neq \nu$, we get the simple results that if $\mu > \nu$, then $p_A^{(n)} \to 0$ when $n \to \infty$, and if $\mu < \nu$, then $p_B^{(n)} \to 0$ when $n \to \infty$. Therefore, when $\mu \neq \nu$, there is global convergence toward a one-locus boundary. Therefore, whatever the initial state, the population will evolve toward a stable equilibrium at which the "healthy" allele (A or B) with the higher mutation rate is eliminated.

If the stable equilibrium is at $\hat{p}_A = 0$, because $\mu > \nu$, the equilibrium frequency of b, the recessive allele still segregating can be found from equation (4b), which at equilibrium reads $\hat{w} = 1 - \nu$ or $1 - sx_4^2 = 1 - \nu$. Since $\hat{x}_4 = \hat{q}_b$ when $p_A = 0$, the equilibrium is given by

$$\hat{x}_4 = \hat{q}_b = \sqrt{\nu/s}. \tag{7}$$

Similarly, if $\mu < \nu$, the segregating mutant has the equilibrium frequency

$$\hat{x}_4 = \hat{q}_a = \sqrt{\mu/s}. \tag{8}$$

So at equilibrium, the frequency ($\hat{x}_4^2$) of the deleterious genotype is μ/s or ν/s, whichever is smaller, and the disorder has turned into a monogenic condition.

Equilibria when $\mu = \nu = \tau$

Consider the case where the two mutation rates are identical and equal to τ. Although this situation may be unrealistic, except with a duplicated locus [4], it provides a useful basis for analysis in which the mutation rates are indeed different.

When the mutation rates are equal, equation (6) reveals that the ratio of the gene frequencies, p_A/p_B, remains unchanged. This reflects the fact that the equations (4a) and (4b) are identical, and we then have exactly two equations to determine equilibrium [i.e., equation (4) and the one remaining independent gametic recurrence equation, for instance equation (2a)]. From equation (4) we get as before

$$\hat{x}_4 = \sqrt{\tau/s}, \tag{9}$$

When we insert this into equation (2a) and parameterize with the gene frequencies instead of the gamete frequencies, we get

$$\begin{aligned}[\tau + (1 - \tau)r](\hat{p}_A + \hat{p}_B) - (1 - \tau)r\hat{p}_A\hat{p}_B \\ = [\tau + (1 - \tau)r](1 - \sqrt{\tau/s}),\end{aligned} \tag{10}$$

which defines a hyperbola of equilibria in the (p_A, p_B) plane as shown in figure 1. [The two single gene equilibria are given as the intercepts $(1 - \sqrt{\mu/s}, 0)$ and $(0.1 - \sqrt{\nu/s})$.]

Since we know that the ratio p_A/p_B is preserved, a population initiated at an arbitrary point, (p_A, p_B), is restricted to a line through this point and the origin. Since the origin and the $p_A = 1$ and $p_B = 1$ boundaries are evidently unstable, the population will evolve along a straight line toward the intersection with the equilibrium hyperbola. Any equilibrium on the hyperbola is stable in all directions except along the hyperbola itself. Whenever a population departs from the hyperbola, it returns to it. Once on the curve, the population is free to drift along it. (For a discussion of the genetic drift in this case see reference [3].)

In the present case, it should be noted from equation (9) that the equilibrium

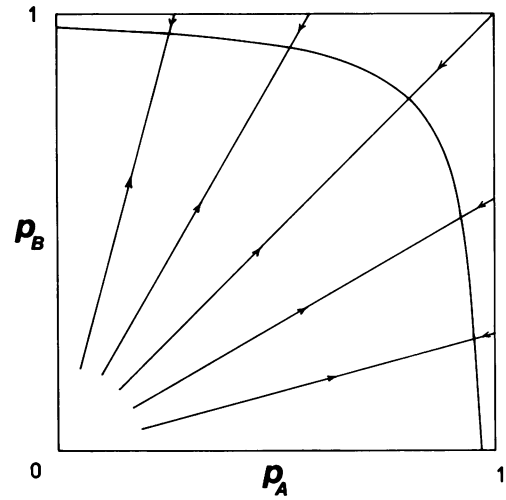

FIG. 1.—The change in gene frequencies through time when the mutation frequencies are equal at the two loci ($\tau = 10^{-3}, r = .1, s = 1$). Trajectory of any population is restricted to a straight line through origin and point of initial gene frequencies. Population will converge to a point on the hyperbola arching between the points $(1 - \sqrt{\tau/s}, 0)$ and $(0, 1 - \sqrt{\tau/s})$.

frequency of the deleterious genotype equals that of a single locus recessive for the whole continuum of equilibria.

The description of the equilibria in terms of the gene frequencies in figure 1 is of course insufficient, but as long as r is large compared to mutation rate τ, the correlation between the two loci is negligible. To see this, consider the linkage disequilibrium at an equilibrium: $\hat{d} = \hat{x}_4 - \hat{q}_a \hat{q}_b$ which from equation (10) may be shown to be $\hat{d} = -\tau \hat{p}_A \hat{p}_B / [\tau + (1 - \tau)r]$, which is always negative. When the mutation rate is assumed small compared to r, we get the approximation

$$\hat{d} = -\tau \hat{p}_A \hat{p}_B / r, \tag{11}$$

which produces an approximate expression for the squared gametic correlation: $\zeta^2 = (\tau/r)^2 (\hat{p}_A \hat{p}_B)/(\hat{q}_a \hat{q}_b)$. From equation (11) the linkage disequilibrium is negative (viz., $\hat{x}_4 < \hat{q}_a \hat{q}_b$). Using this, the gametic correlation may be evaluated as $|\zeta| < (\tau/r)/(\hat{x}_4)^{\frac{1}{2}} = \tau^{(3/4)} s^{(1/4)}/r$. Thus, the gametic correlation between the two loci is less than the square root of the mutation rate.

Dynamics when $\mu \neq \nu$

Consider the equilibrium hyperbola (10) for $\mu = \nu = \tau$, and imagine that ν is increased slightly. This is obviously not going to influence the intercept of the hyperbola with the p_A-axis, while it switches the intercept with the p_B-axis from $(0, 1 - \sqrt{\tau/s})$ to the slightly lower point but higher mutant frequency $(0, 1 - \sqrt{\nu/s})$. At the same time, all points on the perturbed curve become subject to the global convergence (6) toward the intercept with the first axis.

The hyperbola of equilibria (10) in the case of $\mu = \nu = \tau$ may be considered a trajectory of zero velocity between the points $(1 - \sqrt{\mu/s}, 0)$ and $(0, 1 - \sqrt{\nu/s})$. When

ν is increased, this trajectory is deformed slightly; a population on it will move toward the point $(1 - \sqrt{\mu/s}, 0)$. Close to the point $(0, 1 - \sqrt{\nu/s})$, the trajectory between the two one-locus equilibria will be near the hyperbola (10) for $\tau = \nu$, and close to $(1 - \sqrt{\mu/s}, 0)$, it will be near equation (10) for $\tau = \mu$; at any point the trajectory will run between these two hyperbolas. These bounding hyperbolas are indicated by dashed curves in figure 2 in which the trajectory is seen.

The movement along the trajectory between the equilibria on the one-locus boundaries is governed by equation (5). Thus the angular movement in each generation, starting in any point (p_A, p_B) in the plane, is given by

$$\theta' - \theta = \arctan(p_A'/p_B') - \arctan(p_A/p_B) \approx p_A p_B(\mu - \nu)/(p_A^2 + p_B^2), \tag{12}$$

when μ and ν are small. This rate is maximized when $p_A = p_B$, where the rate is $(\mu - \nu)/2$. For any point, the absolute angular rate is indeed small. Reference to figure 2, which shows a number of trajectories calculated for $\nu = 2\mu$, reveals that all the points in the plane are drawn toward the "fundamental trajectory" connecting the equilibrium points on the two axes. The fact that all other trajectories are nearly straight lines except when close to the fundamental trajectory implies that the rate of movement towards it overwhelms the angular movement for all points outside its immediate vicinity.

As when $\mu = \nu$, the trajectories starting at points with high frequency of the deleterious genotype follow very close to the straight line through the origin and the

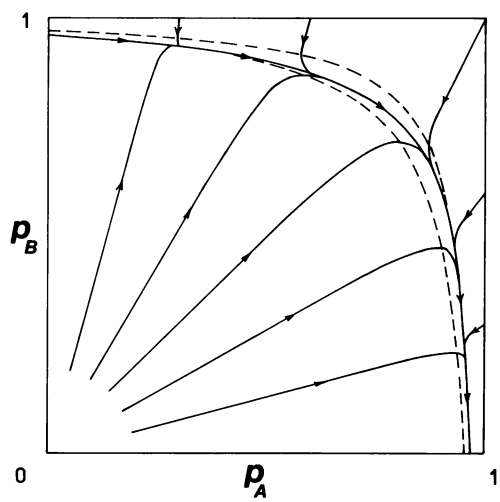

FIG. 2.—The change in gene frequencies through time when the mutation frequencies are different at the two loci ($\nu = 2\mu$, $\mu = 10^{-3}$, $r = .1$, $s = 1$). Population will converge to the globally stable equilibrium $(1 - \sqrt{\mu/s}, 0)$. Population will converge rather fast to fundamental trajectory arching between points $(1 - \sqrt{\mu/s}, 0)$ and $(0, 1 - \sqrt{\nu/s})$, and once on this trajectory, convergence to globally stable equilibrium is very slow. Fundamental trajectory runs between hyperbolas from figure 1 corresponding to $\tau = \mu$ and $\tau = 2\mu = \nu$. These boundaries are shown as dashed curves.

initial point (p_A, p_B). However, starting with the deleterious genotype at very low frequency (i.e., above or to the right of the fundamental trajectory in figure 2), the trajectory is initially very close to a straight line now with the slope

$$\frac{p_A' - p_A}{p_B' - p_B} = \frac{1 - \mu - w}{1 - \nu - w} \frac{p_A}{p_B} \approx \frac{\mu}{\nu} \frac{p_A}{p_B} \tag{13}$$

which has been calculated using equation (4), and where the last approximation is valid for x_4 close to zero. In the case considered above with $\mu < \nu$, the trajectories starting from the $p_A = 0$ or the $p_B = 0$ boundary will be a factor ν/μ steeper than trajectories starting at high values of x_4 (fig. 2).

It can be concluded from the above analysis that a population which starts at an arbitrary point, (p_A,p_B), will be pushed rapidly toward the fundamental trajectory which it will approach with a gene frequency ratio of the "healthy" alleles similar to the initial ratio of p_A:p_B. Once close to or on the fundamental trajectory, further movement, now directed toward the monogenic stable equilibrium, is extremely slow. In fact, as the rate is at most $(\mu - \nu)/2$, it is likely to be negligible compared to random movements in most realistic cases.

RESULTS

Predictions from the Model

If digenic systems complying with our model have emerged in the course of human evolution, a few may have reached their global equilibrium (i.e., entered the realm of ordinary monogenic systems). We are only concerned here with those systems that are still digenic.

The model predicts that these systems will be close to the fundamental trajectory arching toward final global equilibrium. The fundamental trajectory in figure 2 has been drawn for an extremely high mutation rate, $\mu = 10^{-3}$, to make room for the trajectories on the mutation boundaries. For smaller and more realistic mutation rates, the fundamental trajectory almost clings to the two mutation boundaries. Except for a small stretch in the upper right corner, characterized by low mutant frequencies in both loci, all other points on the trajectory have a low mutant frequency in one locus and intermediate gene frequencies in the other. Thus, our model predicts that most of the digenic systems due to balance between mutation and selection will have a typically idiomorphic mutant frequency in one locus and typically polymorphic frequencies in the other locus. In a disorder resulting from such a digenic system, we will observe a locus with a deleterious recessive in mutation-selection equilibrium, where the deleterious genotype has a limited penetrance. The incidence of affected individuals in a family where both parents can produce the gamete *ab* may be deviant from the expected (1/4) with a simple recessive disease.

Incidence of Affected Individuals in Families

Suppose that the alleles *B* and *b* occur at polymorphic frequencies in the population and that the allele *a* occurs at a low frequency corresponding to the mutation selection balance on the fundamental trajectory. The individuals capable of producing the

gamete *ab* are then *AB/ab*, *Ab/aB*, and *Ab/ab*, and they occur approximately in the relative frequencies $p_B/(1 + p_B)$, $p_B/(1 + p_B)$, and $q_b/(1 + p_B)$, respectively. The frequencies with which the three types of individuals produce the gamete *ab* is $(1 - r)/2$, $r/2$, and $1/2$, respectively. Thus, the individuals with the potential of producing *ab* do so with a mean frequency of $1/[2(1 + p_B)]$, such that the incidence of affected individuals from parents which can potentially produce *ab/ab* offspring will be

$$\Pi = [4(1 + p_B)^2]^{-1}. \tag{14}$$

This ratio will be 1/4 at the globally stable equilibrium where $p_B = 0$ and will become 1/16 at the other extreme where p_B is close to unity (table 1).

TABLE 1

FAMILY INCIDENCE AND VARIANCE INCREASE FACTOR

GENE FREQUENCY (p_B)	INCIDENCE	VARIANCE INCREASE FACTOR ϵ	
		$r = 1/2$	$r = 0$
1*	.063	.25	.20
.5	.111	.21	.16
.25	.160	.15	.11
.1	.207	.080	.055
.05	.227	.045	.030
.025	.238	.024	.016
.01	.245	.010	.007
0	.250	0	0

* Interesting only as a limit case.

The different segregation ratios in different families would be hardly detectable in humans. However, the variance in the segregation ratios will be larger than that expected for a binomial distribution with probability parameter Π, namely $\Pi(1 - \Pi)/n$ in a sibship of n. It can be shown that the variance V_n of segregation ratios is increased by a factor $[1 + (n - 1)\epsilon]$ in sibships of n, so

$$V_n = [\Pi(1 - \Pi)/n]\,[1 + (n - 1)\epsilon] \tag{15}$$

where ϵ is a positive function of r and p_B, increasing in p_B and decreasing in r (see Appendix A). Some specific values of the function ϵ are given in table 1. It is apparently rather insensitive to changes in r, whereas it depends heavily on p_B.

The deviation of V_n from the binomial variance increases with n. In humans, however, extensive data cannot be expected for n larger than two, so that the function ϵ gives a good indication of the increase in the observed variance relative to the expected binomial variance. In most cases, the incidence would be observed as deviant from 1/4, and heterogeneity in the segregation ratios would hardly be detected. (For example, with a total sample of 1,000 offspring and using a 5% significance level, a deviation corresponding to $p_B = .05$ would be detected with probability .5, whereas the increased variance among segregation ratios, if all offspring were from sibships of size $n = 2$, would be detected only if p_B were larger than about .2.)

In the calculation of Π and V_n, we have neglected the problem of ascertainment of

families. The probability that a mating capable of producing the deleterious genotype actually does so is very different for the different types of mating. Therefore, in any observation where the matings are ascertained on the basis of the occurrence of the deviant genotype, the matings involving an *Ab/ab* parent will be overrepresented. The effect of this is to push the observed family incidence closer to 1/4 and to decrease the level of heterogeneity, resembling a decrease in p_B with complete ascertainment (see Appendix B). In any case, a deleterious trait with a genetically determined penetrance may give rise to odd segregation ratios without being detected as heterogeneous. However, the exact distribution of segregation ratios for any ascertainment scheme and any p_B may be calculated and compared to an observed distribution.

The Effect of Increased Mutation Rates

The equilibrium population incidence of the disorder is in this model, as with monogenic disorders, proportional to the mutation frequency, as seen from equations (7) and (8) and figure 2. Therefore, the long term change in the frequency of the disorder due to a change in the mutation frequency is the same for a monogenic and a digenic trait.

The largest change in one generation due to a change in the mutation frequency happens right after the change in the mutation frequency. Suppose we have a monogenic disorder determined by the two alleles C and c, with C dominant and the deviant genotype cc having the fitness $1 - s$ relative to 1 for the dominant phenotype and suppose that the allele c is produced from C by mutation in the frequency τ, then at equilibrium we have $\hat{q}_c = \sqrt{\tau/s}$. Suppose now that the mutation frequency has suddenly changed to τ', so that the frequency of c in the following generation is $q_c' = \hat{q}_c + (\tau' - \tau)\,\hat{p}_C/(1 - s\,\hat{q}_c^2)$, then the change in the incidence of the disorder is approximately

$$(q_c')^2 - \hat{q}_c^2 = (q_c' - \hat{q}_c)(q_c' + \hat{q}_c) \approx 2(\tau' - \tau)\,\hat{q}_c. \tag{16}$$

Suppose now that a similar change occurs in our digenic system with equal mutation rates at the two loci (i.e., $\mu = \nu = \tau$), then from equation (2d) we get $x_4' = \hat{x}_4 + (\tau'^2 - \tau^2)\,x_1^* + (\tau' - \tau)(x_2^* + x_3^*)$, which by neglecting terms of the order of squared mutation frequencies reveals the approximate relation

$$x_4' - \hat{x}_4 \approx (\tau' - \tau)(\hat{p}_A\,\hat{q}_b + \hat{q}_a\,\hat{p}_B). \tag{17}$$

If we now assume that the alleles B and b occur in polymorphic frequencies and q_a is small, then equation (17) may be written as

$$x_4' - \hat{x}_4 \approx (\tau' - \tau)\,\hat{q}_b, \tag{18}$$

giving the change in the population incidence corresponding to equation (16) as

$$(x_4')^2 - \hat{x}_4^2 \approx 2(\tau' - \tau)\,\hat{q}_b\hat{x}_4. \tag{19}$$

Thus, the change in the population incidence of the disorder in the digenic system due to an increase in the mutation frequency is less by a factor q_b (the frequency of the modification allele that allow expression) compared to an equivalent monogenic system.

DISCUSSION

The main conclusion from the considerations of this model is that the two-locus disorder will resemble a one-locus disorder, possibly with limited penetrance. This characteristic is shared by a larger class of models where the normal phenotype emerges whenever just one of the loci is homozygous for the normal allele. In particular, it emerges in a situation where the disorder appears to be a one-locus dominant (Christiansen, in preparation).

The model considered here is possibly simplistic. However, consideration of models aimed at the exploration of complex hereditary diseases is important enough to excuse this. Models considering mutation-selection balance in complex genetic systems have been considered by Karlin and McGregor ([5] and references therein) and Wills [6] from the point of view of general evolutionary problems. Wills, in particular, considers a model closely related to ours which provides a mechanism for evolution of sexual isolation between populations.

Our model has two important simplifying assumptions, namely complete recessivity of the gamete *ab* and absence of back mutations from the alleles *a* and *b*. These two assumptions are parallel to the assumptions made in classical treatments of the one-locus recessive model for a hereditary disease, and they give rise to the same kinds of problems originating from the apparent degenerate behavior of the model.

First, the back mutation rate from detrimental alleles to functional alleles is probably at most an order of magnitude less than the rate of production of detrimentals. For the one-locus mutation-selection balance, this rate is too small to have any effect. In the two-locus model the effect is to move the globally stable equilibrium point from the abscissa (fig. 2) by about $\beta/(\nu - \mu)$, where β is the back mutation rate. Since the general conclusion is founded on the observation that the population converges extremely slowly to the equilibrium point, back mutation generally would support our basic thesis.

Secondly, we have to consider the assumption of complete recessivity of the gamete *ab*. As in the one-locus recessive mutation-selection balance model, this assumption is crucial. However, certain types of deviation from the model give qualitatively similar results (Christiansen, in preparation), but in general, we can only expect the description to be valid if the deviations from the recessive model in terms of the selection coefficients are less than the mutation rates. We will refrain from a discussion of the general validity of recessive models and only naively exemplify the model as a pair of regulated enzymes doing exactly the same job and produced in such small quantities that the energy wasted on nonfunctional proteins is negligible. The model can obviously not be related to the duplication series of human hemoglobins, as defective mutants in such a locus most likely show dominance (e.g., like the sickle cell trait).

As argued in the preceding section, the prospect of collecting indirect evidence for the existence of two factor diseases in human populations seems discouraging, and to attempt to ascertain the importance of digenic disorders in human mutational load through analysis of segregation ratios is futile. Therefore, the model used here should be viewed as providing a theoretical example of an irregularly inherited disorder with a genetically determined penetrance. This model has two virtues. First, it points to a

novel explanation of how genetically determined penetrance of rare disorders may emerge, and second, it provides a tool for comparing predictions from single gene disorders to a section of the larger universe of multi-gene disorders.

The change in the incidence of a disease caused by a double recessive condition which follows a rise in mutation frequency has the same functional form as that of a single factor disease. This is important when the hazards of increased radiation exposure of human populations are evaluated. Since the immediate effect of increased mutation frequency on the population incidence is small, recessive genetic disorders may not be the most important object for studies. However, the present study is an introduction to be followed by consideration of digenic disorders caused by alleles with dominance and other kinds of epistatic interactions.

SUMMARY

A simple model for a disease caused by simultaneous occurrence of deleterious recessives at two autosomal loci is suggested and analyzed. Except for the case of equal mutation rates at the two loci, the system will through time converge to a state where the deleterious allele is fixed at the locus with the higher mutation rate, and the system will degenerate to a one-locus mutation-selection balance system. This convergence, however, is very slow, and the system will in the convergence phase resemble a one locus mutation-selection balance system with limited penetrance. The possibility of observing this system in a human population is shown to be limited. The characteristics of the system in a situation of change in mutation rate are discussed.

ACKNOWLEDGMENTS

Part of this work was done while one of the authors (F.B.C.) was visiting the Department of Biological Sciences, Stanford University. The hospitality of this department is gratefully acknowledged. The presentation has benefited from discussions with colleagues at Stanford University and the University of Aarhus. J. S. F. Barker, B. O. Bengtson, and a reviewer made valuable comments on the manuscript.

APPENDIX A

TABLE A1

MEAN SEGREGATION RATIO FOR DIFFERENT MATING COMBINATIONS

Genotype of parent	*AB/ab*	*Ab/aB*	*Ab/ab*
AB/ab	$(1 - r)^2/4$	$r(1 - r)/4$	$(1 - r)/4$
Ab/aB	$r(1 - r)/4$	$r^2/4$	$r/4$
Ab/ab	$(1 - r)/4$	$r/4$	$1/4$
Frequency in population	$p_B/(1 + p_B)$	$p_B/(1 + p_B)$	$q_b/(1 + p_B)$

The calculation of the variance V_n of formula (15) proceeds as follows: V_n = Var (mean segregation ratio given genotypes of parents) + E (variance in segregation ratio given genotypes of parents). The first term is readily calculated from table A1 as

$$[(1-r)^4/16 + 2r^2(1-r)^2/16 + r^4/16]\,p_B{}^2/(1+p_B)^2 + [(1-r)^2/16 + r^2/16]\,2p_Bq_b/(1+p_B)^2 + (1/16)\,q_b{}^2/(1+p_B)^2 - \Pi^2 = \{[(1-r)^2 + r^2]\,p_B/(1+p_B) + q_b/(1+p_B)\}^2/16 - \Pi^2.$$

The second term is the mean of the binomial variances corresponding to the segregation ratios of table A1, thus it becomes

$$\begin{aligned}(1/n)\Big(&\{[(1-r)^2/4][1-(1-r)^2/4] + 2[r(1-r)/4][1-r(1-r)/4] \\ &+ [r^2/4][1-r^2/4]\}\,p_B{}^2/(1+p_B)^2 \\ &+ \{[(1-r)/4][1-(1-r)/4] + [r/4][1-r/4]\}\,2p_Bq_b/(1+p_B)^2 \\ &+ (1/4)[1-(1/4)]\,q_b{}^2/(1+p_B)^2\Big) \\ = \Pi/n &- [1/(16n)]\{[(1-r)^2 + r^2]\,p_B/(1+p_B) + q_b/(1+p_B)\}^2\end{aligned}$$

In total we get

$$\begin{aligned}V_n &= [(n-1)/(16n)]\{[(1-r)^2 + r^2]\,p_B/(1+p_B) + q_b/(1+p_B)\}^2 \\ &\quad + \Pi/n - \Pi^2 \\ &= [\Pi(1-\Pi)/n][1 + (n-1)\epsilon],\end{aligned}$$

where

$$\epsilon = \Big(\{[(1-r)^2 + r^2]\,p_B/(1+p_B) + q_b/(1+p_B)\}^2 - 16\Pi^2\Big)/[16\Pi(1-\Pi)].$$

APPENDIX B

Effect of Limited Ascertainment

The mean incidence of affected individuals is in equation (14) given under the assumption of complete ascertainment of matings which can potentially produce *ab/ab* offspring. To investigate the effect of limited ascertainment we will here discuss a simple model.

Suppose that a mating is always disclosed as segregating *ab/ab* offspring when the first affected offspring occurs. With the segregation proportion α the probability of ascertainment of a sibship of size n is therefore $P_\alpha(n) = 1 - (1-\alpha)^n$. Let us choose the simple procedure of estimating the segregation proportion in ascertained sibships from the offspring following the propositus. Assume further that the number of offspring n from a mating is poisson distributed with mean λ, such that the proportion of ascertained sibships with segregation proportion α is

$$P_\alpha = 1 - \sum_{n=0}^{\infty} (1-\alpha)^n\,[\lambda^n/n!]\exp(-\lambda) = 1 - \exp(-\alpha\lambda).$$

The number of informative offspring (i.e., the number of offspring following propositus) is dependent on the segregation ratio. Designate by n_α this number of offspring, then

$$P(n_\alpha = i \mid n, 0 \le i < n) = \alpha(1-\alpha)^{n-i-1}/[1-(1-\alpha)^n].$$

Thus, the probability of observing a specific n_α from a mating with segregation ratio α becomes

$$P(n_\alpha = i) = \sum_{n=i+1}^{\infty} \alpha(1-\alpha)^{n-i-1}\,(\lambda^n/n!)\exp(-\lambda),$$

and the probability of not observing any is $1 - P_\alpha$. From this the mean observed n_α may be calculated:

$$\begin{aligned} P_\alpha E\, n_\alpha &= \sum_{i=0}^{\infty} \sum_{n=i+1}^{\infty} i\alpha(1-\alpha)^{n-i-1}(\lambda^n/n!)\exp(-\lambda) \\ &= \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} i\alpha(1-\alpha)^{n-i-1}(\lambda^n/n!)\exp(-\lambda) \\ &= \sum_{n=1}^{\infty} \{[n\alpha - 1 + (1-\alpha)^n]/\alpha\}(\lambda^n/n!)\exp(-\lambda) \\ &= \lambda - [1 - \exp(-\alpha\lambda)]/\alpha. \end{aligned}$$

Suppose now that m types of matings exist in the population in the frequencies ϕ_j, $j = 1,2, \ldots, m$, and with the segregation ratios α_j, $j = 1,2, \ldots, m$. If we identify $\bar{n}_j$ with $P_\alpha E\, n_\alpha$ for $\alpha = \alpha_j$, then the expected number of observed offspring from mating type j is proportional to $\bar{n}_j\phi_j$, and the expected number of observed *ab/ab* individuals for mating type j is proportional to $\alpha_j\bar{n}_j\phi_j$. Therefore we expect to observe the mean segregation ratio as

$$\begin{aligned} \Pi^* &= \left(\sum_{j=1}^{m} \alpha_j\bar{n}_j\phi_j\right) / \left(\sum_{j=1}^{m} \bar{n}_j\phi_j\right) \\ &= \left(\sum_{j=1}^{m} \{\alpha_j\lambda - [1 - \exp(-\alpha_j\lambda)]\}\phi_j\right) / \left(\sum_{j=1}^{m} \{\alpha_j\lambda - [1 - \exp(-\alpha_j\lambda)]\}\phi_j/\alpha_j\right) \end{aligned}$$

This segregation ratio may now be calculated and compared with equation (14) by using the segregation ratios of table A1. Table B1 shows the dependence of Π^* on λ for different values of p_B under the assumption that the two loci are unlinked (i.e., $r = 1/2$). The situation $\lambda = \infty$ corresponds to $\Pi^* = \Pi$.

TABLE B1

FAMILY INCIDENCE WITH INCOMPLETE ASCERTAINMENT

GENE FREQUENCY (p_B)	MEAN OFFSPRING NO.			
	2	4	10	∞
.5	.138	.135	.129	.111
.25	.188	.186	.180	.160
.1	.224	.222	.219	.207
.05	.237	.236	.234	.227
.01	.247	.247	.247	.245
0	.250	.250	.250	.250

REFERENCES

1. NEWCOMBE HB: Mutation and the amount of human ill health, in *Radiation Research* edited by NYGAARD OF, ADLER HI, SINCLAIR WK, New York, Academic Press, 1975, pp 937–946

2. FISHER RA: The sheltering of lethals. *Am Nat* 69:446–455, 1935
3. NEI M, ROYCHOUDHURY AK: Probability of fixation of nonfunctional genes at duplicate loci. *Am Nat* 107:362–372, 1973
4. NEI M: Accumulation of nonfunctional genes on sheltered chromosomes. *Am Nat* 104:311–322, 1970
5. KARLIN S, MCGREGOR J: On mutation selection balance for two-locus haploid and diploid populations. *Theor Popul Biol* 2:60–70, 1971
6. WILLS C: A mechanism for rapid allopatric speciation. *Am Nat*. In preparation

Annual Meeting
American Society of Human Genetics

University of California
San Diego, California
October 19–22, 1977

Deadline for Abstracts: Received by June 11, 1977

Send abstracts to:

Dr. Oliver W. Jones, Chairman—Program Committee
Medical Genetics (M-013)
Department of Medicine
University of California, San Diego
La Jolla, California 92093