

POLYMORPHISM AND EVOLUTION OF THE Rh BLOOD GROUPS

Masatoshi NEI, Wen-Hsiung LI, Fumio TAJIMA, and Prem NARAIN¹

*Center for Demographic and Population Genetics, The University of
Texas at Houston, Texas 77425, U.S.A.*

Summary With the aim of understanding the mechanism of maintenance of the Rh polymorphism in man, the probability and the first arrival time of an incompatibility mutant allele (recessive allele r) to reach a high frequency by genetic drift in a finite population and the allele frequency distribution under mutation pressure are studied. The deterministic changes in allele frequency in subdivided populations are also studied. The results obtained are as follows: (1) If the effective population size is 500-1,000, the probability of a single mutant allele to reach a frequency of 0.3 or 0.5 is quite small, and without recurrent mutation it is unlikely that the mutant allele becomes polymorphic. However, if the mutant allele happens to increase in frequency by genetic drift, the increase occurs quite rapidly. (2) In an infinitely large population the backward (u) and forward mutations (v) produce two stable equilibria, one of which has a frequency of 0.065 for $h=0.05$ and a frequency of 0.16 for $h=0.01$ when $u=v=10^{-4}$, where h is the fitness reduction for the offspring from mating $rr \times RR$. These frequencies are substantially higher than 0 but still lower than the frequencies in the European populations (0.3-0.6). In relatively small populations, however, the probability of the allele frequency being 0.3-0.6 becomes quite high if $h=0.01$. (3) If a population is subdivided into subpopulations among which small migration occurs, stable equilibria may be developed. However, the equilibrium gene frequencies do not conform to the frequencies observed in the European populations. When the migration rate becomes higher, the stable equilibria disappear, but the gene frequency change in subdivided populations is generally much slower than that in a single random mating population, so that the Rh polymorphism may be maintained for a long time even if there are no stable equilibria. (4) If we consider all these factors together, it is possible to explain the Rh polymorphism in terms of the mutation-drift hypothesis without recourse to reproductive compensation. It seems that the Rh polymorphism is transient rather than stable.

Received August 25, 1981

¹ On leave from the Indian Institute of Agricultural Statistics, New Delhi, India.

The mechanism of maintenance of the Rh blood group polymorphism in man has been studied by many authors, but the mechanism still remains unknown (see Cavalli-Sforza and Bodmer, 1971 for a review). This blood group system causes mother-fetus incompatibility for certain types of matings and reduces the fitness of heterozygotes. In 1942 Haldane showed that in the absence of any selective forces other than incompatibility selection this polymorphism is unstable and cannot persist for a long time. He therefore proposed that the present polymorphism is caused by relatively recent admixture of populations that differed greatly in gene frequency. His argument was that mutation and random genetic drift can occasionally swing a small population over from one type of homozygosity to the other. Two years later, R.A. Fisher (see Race, 1944) suggested that reproductive compensation by Rh(-) mothers for children lost due to incompatibility might be enough to maintain the Rh polymorphism. This latter hypothesis has subsequently been investigated by many investigators (e.g., Li, 1953; Vogel and Strobel, 1960; Levin, 1967; Feldman *et al.*, 1969). In particular, Feldman *et al.* made a detailed mathematical analysis and concluded that no reasonable degree of compensation or overcompensation can lead to stable equilibria in the observed range of Rh(-) gene frequency. Furthermore, according to Reed (1971) there is no clear evidence of reproductive compensation for Rh(-) mothers. Despite these studies, Haldane's alternative hypothesis has not been well accepted (Feldman *et al.*, 1969). The reason for this seems to be that there are no human populations in which the Rh(-) allele is fixed or nearly fixed, though the Basque population has a frequency of over 0.5. However, it is possible that the high frequencies (0.3-0.6) of the Rh(-) allele in western European populations are caused by random genetic drift, and thus a modified form of Haldane's hypothesis may be valid.

The purpose of this paper is to examine the plausibility of this genetic drift hypothesis with some additional factors. We shall first study the probability and the mean time for the Rh(-) allele to reach a given frequency or become fixed in a population starting from a low frequency, and then examine the gene frequency distribution under mutation pressure. The frequency changes of the Rh(-) allele in subdivided populations will also be examined. The results obtained will be discussed in relation to the polymorphism and evolution of this locus.

FIRST ARRIVAL PROBABILITY AND ARRIVAL TIME

In this section we study two problems by using the diffusion method. One is the first arrival probability, $u(p,y)$, *i.e.*, the probability that the gene frequency reaches a certain frequency y for the first time starting from the initial frequency p , and the other is the mean first arrival time, $\bar{t}(p,y)$, *i.e.*, the mean time required for the gene frequency to reach y starting from p , given that this event will occur. We consider only the D antigen of the Rhesus system since the other alleles are not so important for incompatibility (see *e.g.*, Cavalli-Sforza and Bodmer, 1971). All

Table 1. Mating types and incompatibility selection in the Rh blood groups.

Mating Mother × Father	Frequency	Offspring		
		<i>RR</i>	<i>Rr</i>	<i>rr</i>
<i>RR</i> × <i>RR</i>	$(1-x)^4$	1	0	0
<i>RR</i> × <i>Rr</i>	$2x(1-x)^3$	1/2	1/2	0
<i>Rr</i> × <i>RR</i>	$2x(1-x)^3$	1/2	1/2	0
<i>RR</i> × <i>rr</i>	$x^2(1-x)^2$	0	1	0
<i>rr</i> × <i>RR</i>	$x^2(1-x)^2$	0	1- <i>h</i>	0
<i>Rr</i> × <i>Rr</i>	$4x^2(1-x)^2$	1/4	1/2	1/4
<i>Rr</i> × <i>rr</i>	$2x^3(1-x)$	0	1/2	1/2
<i>rr</i> × <i>Rr</i>	$2x^3(1-x)$	0	(1- <i>k</i>)/2	1/2
<i>rr</i> × <i>rr</i>	x^4	0	0	1

alleles carrying the genetic determinant for *D* will be pooled and denoted by *R*, whereas the other by *r*. This reduces the Rh system to a simple dominant inheritance. We consider a finite population of effective size *N*, and assume that the Hardy-Weinberg law applies. Feldman *et al.* (1969) have shown that this assumption holds approximately.

Table 1 shows the mating types and selection scheme for our model. Our selection scheme is somewhat different from that of Feldman *et al.* (1969), since we do not consider reproductive compensation. The two matings involving incompatibility selection are *rr* mother × *RR* father and *rr* mother × *Rr* father. The proportions of offspring surviving the incompatibility selection are 1-*h* and 1-*k* for the two types of matings, respectively. From this table we can show that the mean fitness of the population is

$$\bar{W} = 1 - x^2(1-x)[h - (h-k)x], \quad (1)$$

and the mean change in gene frequency per generation in diffusion approximations is

$$M_{\delta x} = -\frac{1}{2}x^2(1-x)(1-2x)[h - (h-k)x], \quad (2)$$

whereas the variance is $x(1-x)/(2N)$.

The general formulae for the first arrival probability and the mean first arrival time have been given by Kimura and Ohta (1969, 1973), and Li (1975). They are

$$u(p, y) = g(0, p)/g(0, y) \quad (3)$$

$$\bar{t}(p, y) = \int_p^y \phi(x)[1 - u(x, y)]dx + \frac{1 - u(p, y)}{u(p, y)} \int_0^p \phi(x)u(x, y)dx, \quad (4)$$

where

$$g(a, b) = \int_a^b G(x) dx$$

$$G(x) = \exp \left\{ -4N \int_0^x M_{\delta z} z^{-1} (1-z)^{-1} dz \right\}$$

$$\phi(x) = 4Nu(x, y)g(0, y)/[x(1-x)G(x)].$$

We have used these formulas to carry out some numerical computations. In these computations, we have assumed $k = h/2$ according to the suggestion by Cavalli-Sforza and Bodmer (1971). Examining data on mother-fetus incompatibility, Haldane (1942) and Cavalli-Sforza and Bodmer (1971) have estimated that h is about 0.05. However, the estimate of h varies greatly among reports, some estimates being the order of 0.01 (Walker and Murray, 1956; Cavalli-Sforza and Bodmer,

Table 2. First arrival probabilities for allele r to reach frequency y starting from frequency p . k is assumed to be $h/2$.

p	h	N	y				
			0.10	0.30	0.50	0.70	1.00
0.10	0.05	50	1	0.320	0.183	0.129	0.096
		100	1	0.306	0.167	0.115	0.090
		500	1	0.204	0.070	0.043	0.040
		1,000	1	0.108	0.018	0.010	0.010
	0.01	50	1	0.331	0.197	0.140	0.099
		500	1	0.306	0.167	0.115	0.090
		1,000	1	0.279	0.137	0.091	0.077
	0	any size	1	0.333	0.200	0.143	0.100
		0	any size	1	0.333	0.200	0.143
	1/2N	0.05	50	0.099	0.032	0.018	0.013
100			0.049	0.015	0.008	0.006	0.004
500			0.009	0.002	6.5×10^{-4}	4.0×10^{-4}	3.7×10^{-4}
1,000			0.004	4.6×10^{-4}	7.8×10^{-5}	4.4×10^{-5}	4.3×10^{-5}
0.01		50	0.100	0.033	0.020	0.014	0.010
		100	0.050	0.016	0.010	0.007	0.005
		500	0.010	0.003	0.002	0.001	8.8×10^{-4}
		1,000	0.005	0.001	6.7×10^{-4}	4.4×10^{-4}	3.7×10^{-4}
0		50	0.100	0.033	0.020	0.014	0.010
		100	0.050	0.017	0.010	0.007	0.005
		500	0.010	0.003	0.002	0.001	0.001
		1,000	0.005	0.002	0.001	7.0×10^{-4}	5.0×10^{-4}

1971). Particularly we note that blood transfusions for Rh(-) mothers before their pregnancy often give an overestimate of h . We have, therefore, used $h=0.05$ and 0.01 . The case of complete neutrality ($h=k=0$) is also included for comparison.

First arrival probability. The results for first arrival probabilities are given in Table 2. In the case of neutral genes $u(p,y)$ is given by p/y , and thus it is independent of population size and decreases as y increases. When $h>0$, $u(p,y)$ is smaller than that for neutral genes, as expected. This can be seen from the case of $p=0.1$. If $h=0.01$ and the population size (N) is 1,000 or smaller, $u(p,y)$ is only slightly smaller than that for neutral genes. Namely, in this case the incompatibility genes behave as though they are neutral. This is true even for $h=0.05$ as long as N is 100 or smaller. However, as N increases, $u(p,y)$ rapidly decreases in this case. Thus, in the case of $h=0.05$ and $N=1,000$, $u(0.1, 0.5)$ is about 1/11 of that for neutral genes.

All mutant genes initially exist as a single copy in the population. Therefore, the case of $p=1/2N$ is of special interest. In this case $u(p,y)$ is $1/(2Ny)$ for neutral genes and obviously decreases as N increases. As expected, incompatibility selection reduces $u(p,y)$, but if $h=0.01$, the effect of selection is negligibly small when $N\leq 500$. Only when N is as large as 1,000, $u(p,y)$ is reduced appreciably, compared with that for neutral genes. When $h=0.05$ and $N=500$, $u(1/2N, 0.5)$ is about 1/3 of that for $h=0$, but the values of $u(p,y)$ for $y=0.7$ and $y=1.0$ are nearly the same as that for $y=0.5$. In the case of $N=1,000$, $u(p,y)$ is further reduced, but the values of $u(1/2N, 0.7)$ and $u(1/2N, 1.0)$ are again nearly the same as that of $u(1/2N, 0.5)$. The reason for this is that in the presence of sufficiently strong incompatibility selection allele r will almost always become fixed in the population once the gene frequency passes 0.5.

The first arrival probability for a single mutation to reach a frequency of 0.3 or 0.5 is very small in a large population. In practice, however, the mutation from allele R to r may occur recurrently. If this is the case, the frequency of allele r can be substantial even in large populations. However, this problem can be studied more appropriately by examining the gene frequency distribution.

In the above study we assumed that allele r was introduced later than allele R . In practice we do not know which allele was the first. It is, therefore, interesting to know the probability that allele R rather than r reaches a polymorphic status from a low frequency. This probability was computed for the case of $N=500$ and $p=1/2N$. If we assume $h=0.05$ and $k=0.025$, this probability becomes 7.6×10^{-5} , 1.3×10^{-5} , 7.2×10^{-6} , and 5.8×10^{-6} for $y=0.3$, 0.5 , 0.7 , and 1.0 , respectively. If we compare these values with those in Table 2, it is clear that the probability of allele R becoming polymorphic is much smaller than that of allele r . For example, in the case of $y=1.0$ the former is about 1/10,000 of the latter. Therefore, in this case it is not easy to "swing a population over from the homozygosis for allele r to the other homozygosis." If the h value is as small as 0.01, the probability of the frequency of R reaching a high frequency increases, but yet it is very small. For example, in the case of $N=500$, $p=1/2N$, and $y=1.0$, it is 3.7×10^{-4} , which is about 1/200 of the probability of fixation of allele r . This makes Haldane's hypothesis

Table 3. Mean first arrival time in units of $4N$ generations for allele r to reach frequency y starting from frequency p . k is assumed to be $h/2$.

p	h	N	y					
			0.10	0.30	0.50	0.70	1.00	
0.10	0.05	50	0	0.117	0.255	0.423	0.868	
		100	0	0.117	0.255	0.413	0.800	
		500	0	0.120	0.238	0.340	0.513	
		1,000	0	0.115	0.200	0.266	0.367	
	0.01	50	0	0.116	0.255	0.430	0.993	
		500	0	0.117	0.255	0.413	0.800	
		1,000	0	0.119	0.253	0.395	0.696	
		any size	0	0.116	0.255	0.432	0.948	
	$1/2N$	0.05	50	0.047	0.164	0.302	0.470	0.914
			100	0.049	0.167	0.304	0.463	0.850
			500	0.052	0.171	0.290	0.392	0.565
			1,000	0.053	0.168	0.253	0.319	0.419
0.01		50	0.047	0.163	0.302	0.477	0.977	
		100	0.049	0.166	0.304	0.478	0.963	
		500	0.051	0.169	0.306	0.465	0.852	
		1,000	0.052	0.170	0.304	0.446	0.747	
0		50	0.047	0.163	0.302	0.479	0.995	
		100	0.049	0.165	0.304	0.482	0.997	
		500	0.051	0.167	0.306	0.484	1.000	
		1,000	0.052	0.168	0.307	0.484	1.000	

of racial mixture less tenable, but we cannot rule out the possibility that racial admixture occurred as a rare event right after allele r was fixed in a population. Moreover, recurrent mutation may occur between the two alleles, and in this case, the plausibility of Haldane's hypothesis increases, as will be discussed later.

First arrival time. The mean first arrival times $[\bar{i}(p,y)]$ for the same sets of genetic parameters as those for Table 2 are given in Table 3. For neutral genes we have

$$\bar{i}(p,y) = 4N \left[\frac{1-y}{y} \ln(1-y) - \frac{1-p}{p} \ln(1-p) \right]$$

(Kimura and Ohta, 1973). Therefore, it is proportional to $4N$. In Table 3 the mean first arrival time is given in units of $4N$ generations, so that the values for neutral genes for given values of p and y are applicable to any population size. For example, in the case of $p=0.1$ and $y=0.3$, $\bar{i}(p,y)=0.116 \times 4N$ generations. Thus, it is 23.2 generations for $N=50$ but 232 generations for $N=500$. In the presence of in-

compatibility selection $\bar{t}(p,y)$ is a complicated function of p , y , and N . In the case of $p=0.1$ the effect of selection is small for $y=0.5$ as long as N is 500 or less. As N increases further, however, the ratio of $\bar{t}(p,y)$ to that for neutral genes tends to decrease. (Strictly speaking, as N increases, this ratio first increases slightly and then declines.) The reason for this is that as N increases, the arrival probability decreases, so that for the gene frequency to reach y it has to do so quickly.

The effect of incompatibility selection on the mean first arrival time for the case of $p=1/2N$ is essentially the same as that for $p=0.1$; the effect is small unless population size is large. In the case of $h=0.05$ and $y=0.1$, the ratio of $\bar{t}(p,y)$ to that for neutral genes increases as N increases from 50 to 1,000, but as N further increases it starts to decrease though it is not shown here. The same is true for the case of $h=0.01$ and $y=0.3$. At any rate, it is clear from this study that the polymorphism of incompatibility genes is established quite rapidly in small populations, though the probability of the establishment is rather small. For example, if $N=500$ with $h=0.01$, we have $\bar{t}(1/2N, 0.3)=0.169$. This corresponds to 338 generations or 7,000–10,000 years since one generation in the past was probably between 20 to 30 years. If $N=1,000$, the mean first arrival time becomes 14,000–20,000 years. This is still a relatively short time in human evolution.

Although the mean first arrival time gives a rough idea about how quickly the frequency of a mutant allele increases in a population, a more complete picture is obtained by examining the distribution of first arrival time. We examined this distribution by using the usual Markov chain method and iterating the gene frequency transition matrix (e.g., Nei, 1975, p. 82). The results obtained for the case of $h=0.05$, $p=1/2N$, and $N=50$ are given in Fig. 1. It is clear that the distribution is quite similar to the gamma distribution and highly leptokurtic. Furthermore,

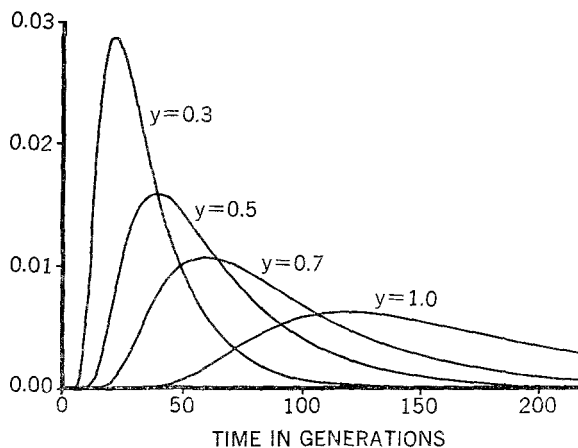


Fig. 1. Distributions of first arrival times for the frequency of allele r to reach y from $1/2N$ in a population of effective size (N) equal to 50. These distributions are conditional, since only the events in which the allele frequency reaches y are considered. $h=0.05$; $k=0.025$.

when y is large, the variance of first arrival time is very large. Therefore, there is a substantially high probability that the arrival to a given frequency occurs earlier than the mean arrival time.

The means (variances) of first arrival times obtained by the present method for $y=0.3, 0.5, 0.7$ and 1.0 are 35.6 (429), 64.1 (1,364), 96.1 (2,960), and 180.7 (8,528) generations, respectively. The corresponding values from Table 3 are 32.8, 60.4, 94.0 and 182.8 generations, respectively. In the case of $y=1$ the diffusion method gives a somewhat larger value than the Markov chain method, as is generally the case (Ewens, 1963; Narain, 1970). When $y < 1$, however, the reverse is true. The reason for this is that in the Markov chain method the change in gene frequency is discrete, so that in the case of $y < 1$ the gene frequency has often reached a value higher than y when the arrival time is recorded.

EFFECT OF MUTATION

Let us now consider the effect of mutation on the Rh polymorphism. At this locus many alleles have been discovered in man and his related species chimpanzees, and it is possible that the forward and backward mutations are occurring between recessive and dominant alleles, though all the recessive or all the dominant alleles may not be the same at the nucleotide level. If this is the case and the population size stays constant, the frequency of allele r is expected to follow a certain frequency distribution. If we denote the mutation rate per generation from R to r by u and the backward mutation rate by v , this distribution is given by the following formula.

$$\phi(x) = C e^{2N\bar{W}} x^{4Nu-1} (1-x)^{4Nv-1}, \quad (5)$$

where C is the normalizing factor and \bar{W} is the mean fitness of the population given by (1) (Wright, 1969).

Before studying the property of this distribution, let us examine the deterministic effects of mutation and selection. Haldane (1944) and Vogel (1954) studied the equilibrium gene frequencies when there is no backward mutation, but in the presence of backward mutation the property of equilibrium gene frequencies changes substantially. In this case the expected frequency change per generation is given by

$$\Delta x = (1-x)u - xv - \frac{1}{2}x^2(1-x)(1-2x)[h - (h-k)x] \quad (6)$$

approximately. We consider the special case of $u=v$ and $k=h/2$ in this paper. In this case (6) reduces to

$$\Delta x = (1-2x) \left[u - \frac{h}{4}x^2(1-x)(2-x) \right]. \quad (6a)$$

The usual mutation rate used in population genetics is 10^{-5} , but in the case of Rh blood groups it could be higher than this, because this polymorphism is appar-

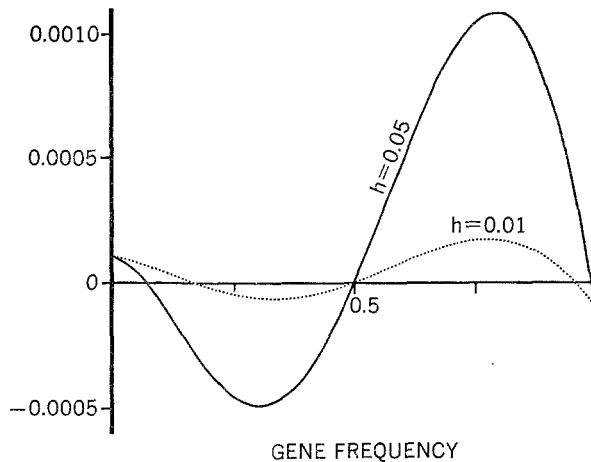


Fig. 2. Relationships between the amount of change in allele frequency (ordinate) and the allele frequency (abscissa) under mutation pressure. These relationships were obtained by formula (6a).

ently controlled by complex loci. Figure 2 shows the value of Δx for the case of $u=10^{-4}$. It is clear that in both cases of $h=0.05$ and $h=0.01$ there are an unstable equilibrium (at $x=1/2$) and two stable equilibria. (When h is sufficiently large, the two stable equilibria are given by $\sqrt{2u/h}$ and $1-4u/h$ approximately.) In the case of $h=0.05$, the lower and higher stable equilibrium points are $x=0.065$ and $x=0.992$, respectively. These values are quite different from the gene frequency for the European populations. If $h=0.01$, the equilibrium frequencies become 0.16 and 0.96, but they are still far from the European frequency. Therefore, in order to explain the present polymorphism in terms of mutation-selection balance we must assume an unrealistically high mutation rate (Cavalli-Sforza and Bodmer, 1971). The story, however, changes if we consider the effect of random genetic drift as will be seen below.

Let us now examine the gene frequency distribution determined by (5). Since it is difficult to find an explicit expression of C in (5), we have used a numerical method to study the distribution. Figure 3 shows the distributions for the cases of $h=0.05$ and 0.01 with $u=v=10^{-4}$ and $N=10^3$. It is clear that the distributions are all U-shaped, and the probability of intermediate frequency near 0.5 decreases as h increases. With $h=0.05$ and 0.01 the distribution is U-shaped as long as $4Nu$ is smaller than 1. If $4Nu$ becomes larger than 1, the distribution for neutral genes tends to be bell-shaped, whereas that for allele r tends to be bimodal, the two peaks being near the two stable equilibrium frequencies under the deterministic treatment. In this paper we are particularly interested in explaining the Rh polymorphism in the European populations, so that we examined the probabilities of the gene frequency being in various frequency intervals. Table 4 shows these probabilities for

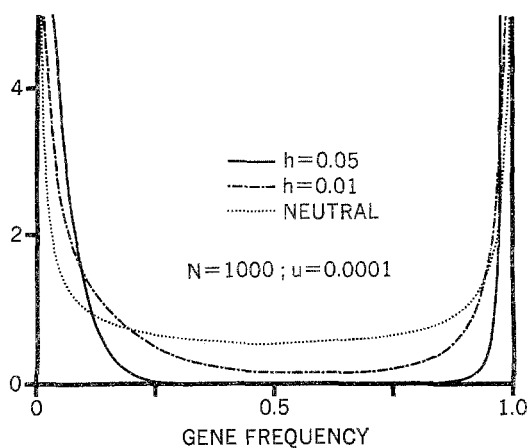


Fig. 3. Distributions of the frequency of allele r in a population of $N=1,000$. The forward (u) and backward (v) mutation rates are assumed to be 10^{-4} .

Table 4. Probabilities of the frequency of allele r being within certain intervals. k is assumed to be $h/2$.

u	v	h	N	Frequency class						
				0.00-0.05	0.05-0.30	0.30-0.60	0.60-0.95	0.95-1.00		
10^{-5}	10^{-5}	0.05	500	0.500	0.010	3×10^{-4}	0.002	0.487		
			10^3	0.500	0.013	2×10^{-5}	7×10^{-4}	0.487		
			10^4	0.720	0.007	7×10^{-26}	9×10^{-13}	0.273		
		0.01	500	0.477	0.017	0.005	0.013	0.488		
			10^3	0.466	0.029	0.005	0.015	0.485		
			10^4	0.581	0.098	9×10^{-7}	5×10^{-4}	0.321		
		10^{-4}	10^{-4}	0.05	500	0.435	0.100	0.003	0.019	0.442
					10^3	0.480	0.153	3×10^{-4}	8×10^{-3}	0.359
					10^4	0.668	0.332	4×10^{-22}	3×10^{-11}	2×10^{-4}
0.01	500			0.331	0.135	0.050	0.107	0.378		
	10^3			0.293	0.233	0.055	0.123	0.296		
	10^4			0.074	0.920	4×10^{-4}	0.002	0.003		
10^{-4}	10^{-5}			0.05	500	0.055	0.013	4×10^{-4}	0.004	0.927
					10^3	0.033	0.011	3×10^{-5}	0.001	0.955
					10^4	2×10^{-6}	1×10^{-6}	4×10^{-26}	3×10^{-12}	0.999
		0.01	500	0.051	0.021	0.009	0.024	0.895		
			10^3	0.030	0.025	0.007	0.027	0.911		
			10^4	3×10^{-6}	5×10^{-5}	5×10^{-8}	0.001	0.999		

different values of u , v , h , and N . When $u=v=10^{-5}$ and $h=0.05$, the gene frequency is either 0 or 1 in most of the time if $N \leq 10^3$. As N increases, the gene frequency reaches one of the two deterministic equilibria. However, since these equilibrium frequencies are far from the European frequencies, we can conclude that the probability of polymorphism, particularly, the probability of gene frequency being in the range of 0.3–0.6 [$P(0.3-0.6)$], is very small. When $h=0.01$ and $u=v=10^{-5}$, the probability of polymorphism [$P(0.05-0.95)$] increases but $P(0.3-0.6)$ is still less than 1 percent. However, if the mutation rate increases, the latter probability increases substantially. For example, in the case of $u=v=10^{-4}$ and $h=0.01$, $P(0.3-0.6)$ is 0.055 for $N=10^3$. Furthermore, if we consider the gene frequency range of 0.3–0.95, the probability becomes 0.178.

The above results indicate that in order to explain the Rh polymorphism by mutation and genetic drift a small value of h , a relatively high mutation rate, and a relatively small population size are required. We have already indicated the possibility that h is as small as 0.01. The mutation rate could also be as high as 10^{-4} for the Rh blood group systems. We also note that before agriculture was introduced the effective size of human populations was probably so small that $N=10^3$ may not be unrealistic at all. (The effective size of a human population is probably about 1/3 of the census population size.) The current effective size of many European populations is certainly much larger than 10^3 , but once the gene frequency reaches around 0.5, it takes a long time for the frequency to decrease or increase in large populations. For example, if the initial frequency is 0.48, the gene frequency declines to 0.127 in 1,000 generations if $h=0.05$ and to 0.451 if $h=0.01$. This suggests that a modified form of Haldane's hypothesis is still tenable.

POLYMORPHISM IN SUBDIVIDED POPULATIONS

So far we have considered a single random mating population. In practice, however, human populations are often divided into subpopulations among which some migration occurs. In these subdivided populations it is possible to maintain the Rh polymorphism if allele r is higher than 0.5 in some subpopulations and less than 0.5 in others, since in the two groups of subpopulations different alleles are advantageous. This idea is interesting, since the Basque population has a frequency of about 0.6, whereas in the neighboring populations the frequency is lower than 0.5. Karlin and McGregor's (1972) small-parameter theory indicates that there are always stable equilibria if the migration rate is small. However, when the migration rate is high, their theory cannot be used. We have therefore studied this problem in some detail by using a deterministic theory.

For simplicity we consider two populations and denote by m_1 and m_2 the proportions of individuals exchanged in populations 1 and 2, respectively. Namely, if the sizes of populations 1 and 2 are N_1 and N_2 , respectively, and the number of individuals exchanged per generation is M , $m_1=M/N_1$ and $m_2=M/N_2$. Therefore,

$m_1 = m_2$ if $N_1 = N_2$ but $m_1 \neq m_2$ otherwise. Let x and y be the frequencies of r in populations 1 and 2, respectively, and assume that selection occurs before migration. The gene frequencies in the following generations (x', y') are then given by

$$x' = (1 - m_1)\xi + m_1\zeta \quad (7a)$$

$$y' = (1 - m_2)\zeta + m_2\xi \quad (7b)$$

approximately, where

$$\xi = \left[x - \frac{1}{2} x^2(1-x)\{h - (h-k)x\} \right] / [1 - x^2(1-x)\{h - (h-k)x\}],$$

$$\zeta = \left[y - \frac{1}{2} y^2(1-y)\{h - (h-k)y\} \right] / [1 - y^2(1-y)\{h - (h-k)y\}].$$

The above formulae are approximate, because we have assumed the Hardy-Weinberg equilibrium in each population. In subdivided populations the deviation of genotype frequencies from Hardy-Weinberg proportions is greater than that in a single population. However, our numerical computations have shown that the actual gene frequency changes when the assumption of Hardy-Weinberg equilibrium is removed are close to those given by (7a) and (7b). Therefore, we shall use these formulae.

The equilibrium frequencies can be determined by equating x' to x in (7), but the analytical solution is not simple. Therefore, we examined the amounts of changes in x and y per generation ($\Delta x, \Delta y$) for various values of x and y . In practice, we divided the gene frequency range (0-1) into 100 equal intervals and computed Δx and Δy for all combinations of 100 x values and 100 y values. This enabled us to determine the trajectories of gene frequency changes, equilibria, and the stability of the equilibria for various values of m_1 , m_2 , and h . ($k = h/2$ was assumed.) The results of this study can be summarized as follows:

(1) The point (0.5, 0.5) is always an unstable equilibrium, as expected. When $m_1 = m_2$, the gene frequency changes are symmetrical around the line between (0, 0) and (1, 1). This symmetry disappears when $m_1 \neq m_2$.

(2) If $h \leq 0.05$, stable equilibria occur only when m_1 and m_2 are very small. When $m_1 = m_2 = 0.0005$ and $h = 0.05$, there are two (nontrivial) stable equilibria, *i.e.* $(x, y) = (0.2, 0.9)$ and $(0.9, 0.2)$, and five unstable equilibria (Fig. 4a). The stable equilibrium gene frequencies observed in this case, however, do not agree with the frequencies in the Basque and its surrounding populations. The results for smaller values of h are qualitatively the same as those for $h = 0.05$, but the stability of the equilibria is generally weaker. In the following we therefore consider only the case of $h = 0.05$.

(3) When $m_1 = m_2 = 0.002$, there are no stable equilibria but there exist three unstable equilibria (Fig. 4b). These three unstable equilibria are relatively closely located, and in the region of these equilibria the gene frequency changes are very

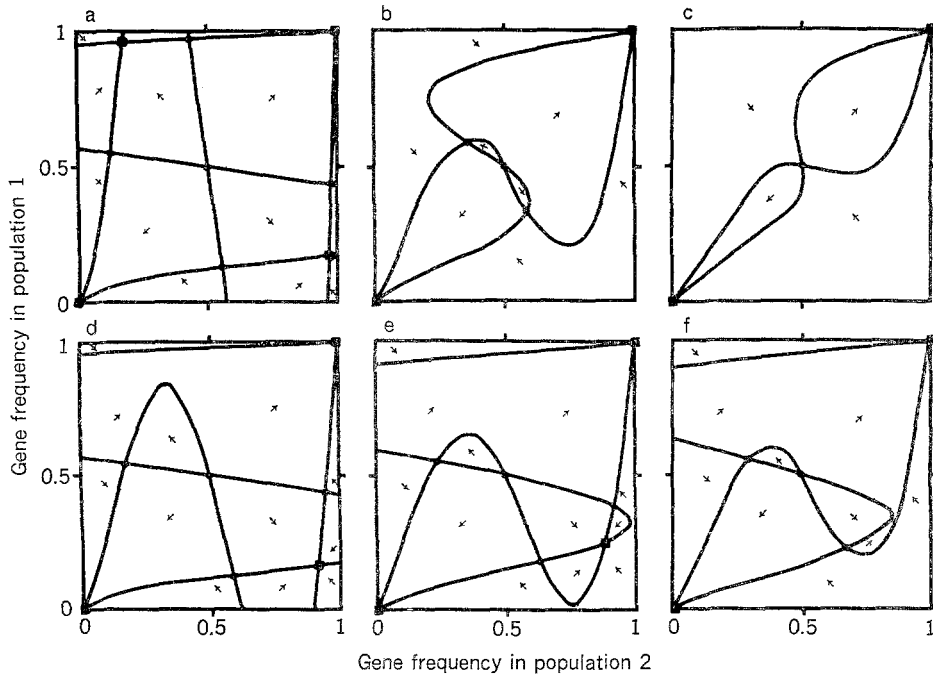


Fig. 4. Allele frequency trajectories and equilibria in a population, which is subdivided into two, populations 1 and 2. The proportion of individuals exchanged per generation is m_1 in population 1 and m_2 in population 2. ■ and ● represent stable and unstable equilibria, respectively. The arrow signs show the direction of allele frequency changes. (a) $m_1=m_2=0.0005$. (b) $m_1=m_2=0.002$. (c) $m_1=m_2=0.004$. (d) $m_1=0.0005$; $m_2=0.001$. (e) $m_1=0.0008$; $m_2=0.0015$. (f) $m_1=0.001$; $m_2=0.002$.

slow. Therefore, even if these equilibria are unstable, they have an effect to maintain the polymorphism for a long time if proper initial gene frequencies are given. For example, if the initial gene frequencies are (0.48, 0.515), the gene frequencies after 1,000 generations are still (0.329, 0.364). In the case of a single random mating population the gene frequency changes from 0.48 to 0.127 in 1,000 generations. We note that 1,000 generations corresponds to 20,000–30,000 years in man.

(4) If $m_1 (=m_2)$ further increases, the unstable equilibria except the one at (0.5, 0.5) gradually disappear (Fig. 4c). After disappearance of these equilibria, (0.5, 0.5) becomes a saddle point.

(5) When $m_1 \neq m_2$, the number of stable equilibria may not necessarily be two even if m_1 and m_2 are small. In the case of $m_1=0.0005$ and $m_2=0.001$ only one stable equilibrium (0.18, 0.92) and four unstable equilibria exist (Fig. 4d). The stable equilibrium gene frequency in the smaller population (with a larger migration rate) is larger than that in the larger population. This is interesting, because the gene frequency in the Basque population whose size is small is higher than that in the neighboring population. However, the equilibrium frequencies do not coincide

with those of the Basque and its neighboring populations. The equilibrium gene frequencies (0.24, 0.82) for the case of $m_1=0.0008$ and $m_2=0.0015$ (Fig. 4e) are closer to the observed frequencies (0.4, 0.6), but the discrepancy is still substantial.

(6) If we increase m_1 and m_2 , the stable equilibrium again disappears, and the unstable equilibria tend to be located closely (Fig. 4f). The trajectories of gene frequency changes are also complicated, and despite the absence of stable equilibrium, the gene frequencies may stay polymorphic for a long time. Thus, if $m_1=0.001$, $m_2=0.002$, and the initial frequencies are (0.48, 0.53), the frequencies after 1,000 generations are still (0.326, 0.534) and those after 1,500 generations are (0.166, 0.236).

The above results indicate that subdivision of populations alone cannot explain the current Rh polymorphism in European populations but certainly helps in prolonging the polymorphism.

DISCUSSION

The present study suggests that in the presence of recurrent mutation the Rh polymorphism in man can be explained by random genetic drift. We have seen that if the effective population size is as small as 1,000 and $h=0.01$ and $u=v=10^{-4}$, the probability of gene frequency lying between 0.3 and 0.6 [$P(0.3-0.6)$] is about 6 percent. Therefore, if we consider that the effective sizes of human populations have increased only recently, it is possible to explain the Rh polymorphism in man in terms of mutation and genetic drift. However, if h is as high as 0.05 and the mutation rate is lower than 10^{-4} , $P(0.3-0.6)$ is quite small. Thus, if the latter parameter values are correct, we must conclude that the Rh polymorphism has occurred as a relatively rare event. It is therefore important to know more about h , u , and v .

In the long history of human evolution, however, even a relatively rare event cannot be neglected, since similar events might have occurred many times. We have noted that although the allele frequency of 0.5 is unstable, the Rh polymorphism may be maintained for a long time particularly in subdivided populations. If we consider this possibility, it is not difficult to explain the polymorphism in terms of mutation and genetic drift even if h is as large as 0.05. One possible scheme of the evolution of the Rh polymorphism in man is as follows: At the time of the last glaciation (Wurm-Wisconsin Period; about 10,000 to 60,000 years ago) there were probably many isolated small human populations. In these populations the frequency of r was generally kept in low frequency by the balance of mutation, incompatibility selection, and genetic drift. In a few populations, however, the allele frequency increased substantially by genetic drift, reaching a frequency higher than 0.5. In Table 2 we note that if the initial frequency is 0.1, the probability of the allele frequency to increase to more than 0.5 is substantial if $N=50-500$. Probably this event occurred in the ancestral population of the present western Europeans. When the glaciation ended, some populations started to increase in size. Probably the ancestral population of Europeans was one of those populations. Admixture

with other populations is likely to have reduced the allele frequency of r in this population to the level close to 0.5, but since the allele frequency change per generation is very small when the frequency is near 0.5, the polymorphism has persisted for a long time. We have seen that the Rh polymorphism can persist for more than 1,000 generations in subdivided populations, and the period of 1,000 generations in man is longer than the post-glaciation period of 10,000 years. We believe that this scheme of evolution is quite likely to have happened. One supporting evidence is that the frequency of r is the highest in the western European populations and declines as the distance from Europe increases (Mourant *et al.*, 1976). In Asian, Amerindian or Oceanian populations the frequency is quite low.

In the past whenever a new genetic polymorphism was discovered, population geneticists looked for the mechanism of maintenance for the polymorphism, assuming that it is stable. Recent molecular data, however, indicate that nucleotide or codon substitution in genes has occurred continuously in the evolutionary process and thus there must be many transient polymorphisms. As shown in this paper, it is quite possible that the Rh polymorphism is one such example. In our study we considered the forward and backward mutations between the group of dominant alleles and the group of recessive alleles. At the nucleotide level, however, every allele could be unique, though the alleles are classified into two groups in terms of the phenotypic effect. Therefore, our model does not contradict the notion that gene substitution is continuously occurring at the Rh locus. The idea of transient polymorphism of the Rh locus is supported by the fact that the number of "alleles" shared by man, chimpanzee, gorilla, and orangutan are very few (Socha and Moor-Jankowski, 1979).

In this paper we did not consider such a stabilizing selection as overdominance. If there is heterozygote advantage in viability, a stable polymorphism is of course maintained, as shown by Feldman *et al.* (1969). However, there is no evidence for this in the Rh blood group system.

Acknowledgement This work was supported by research grants NIH GM 20293, NSF DEB-76-06069, and NSF DEB-77-09120.

REFERENCES

- Cavalli-Sforza, L.L., and Bodmer, W.F. 1971. *The Genetics of Human Populations*. W.H. Freeman, San Francisco.
- Ewens, W.J. 1963. Numerical results and diffusion approximation in a genetic process. *Biometrics* **50**: 241-249.
- Feldman, M.W., Nabholz, M., and Bodmer, W.F. 1969. Evolution of the Rh polymorphism: A model for the interaction of incompatibility, reproductive compensation, and heterozygote advantage. *Am. J. Hum. Genet.* **21**: 171-193.
- Haldane, J.B.S. 1942. Selection against heterozygosis in man. *Ann. Eugenics* **11**: 333-340.
- Haldane, J.B.S. 1944. Mutation and the Rhesus reaction. *Nature* **153**: 106.
- Karlin, S., and McGregor, J. 1972. Application of method of small parameters to multiniche population genetic models. *Theoret. Pop. Biol.* **3**: 186-209.

- Kimura, M., and Ohta, T. 1969. The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**: 763-771.
- Kimura, M., and Ohta, T. 1973. The age of a neutral mutant persisting in a finite population. *Genetics* **75**: 199-212.
- Levin, B.R. 1967. The effect of reproductive compensation on the long term maintenance of the Rh-polymorphism: the Rh crossroad revisited. *Am. J. Hum. Genet.* **19**: 288-302.
- Li, C.C. 1953. Is the Rh facing a crossroad? A critique of the compensation effect. *Am. Natl.* **87**: 257-261.
- Li, W.-H. 1975. The first arrival time and mean age of a deleterious mutant gene in a finite population. *Am. J. Hum. Genet.* **27**: 274-286.
- Mourant, A.E., Kopeć, A.C., and Domaniewska-Sobczak, K. 1976. *The Distribution of the Human Blood Groups and Other Polymorphisms*, 2nd ed. Oxford University Press, London.
- Narain, P. 1970. A note on the diffusion approximations for the variance of the numbers of generations until fixation of a neutral mutant gene. *Genet. Res.* **15**: 251-255.
- Nei, M. 1975. *Molecular Population Genetics and Evolution*. North Holland, Amsterdam and New York.
- Race, R.R. 1944. Some recent observations on the inheritance of blood groups. *Brit. Med. Bull.* **2**: 160-165.
- Reed, T.E. 1971. Dogma disputed: Does reproductive compensation exist? An analysis of Rh data. *Am. J. Hum. Genet.* **23**: 215-224.
- Socha, W.W., and Moor-Jankowski, J. 1979. Blood groups of anthropoid apes and their relationship to human blood groups. *J. Hum. Evol.* **8**: 453-465.
- Vogel, F. 1954. The mutation rate of the Rh-loci—a critical review. *Am. J. Hum. Genet.* **6**: 279-283.
- Vogel, F., and Strobel, D. 1960. Über die Populationsgenetik der ABO-Blutgruppen. 1. Mitteilung. Mathematische Blutgruppen-Selektionsmodelle. *Acta Genet. Statist. Med.* (Basel) **10**, 247-267.
- Walker, W., and Murray, S. 1956. Haemolytic disease of the newborn as a family problem. *Brit. Med. J.* **12**: 187-193.
- Wright, S. 1969. *Evolution and the Genetics of Populations. Vol. 2. The Theory of Gene Frequencies*. University of Chicago Press, Chicago.