Blood Cells, Molecules, and Diseases (2000) **26**(4) August: 348–359
doi:10.1006/bcmd.2000.0314, available online at http://www.idealibrary.com on IDEAL®

Linkage to Gaucher Mutations in the Ashkenazi Population: Effect of Drift on Decay of Linkage Disequilibrium and Evidence for Heterozygote Selection

Submitted 07/07/00
(Communicated by Ernest Beutler, M.D., 07/07/00)

F. Edward Boas[1,2]

ABSTRACT: The two most common Gaucher disease mutations in the Ashkenazi population, 1226A→G and 84G→GG in the glucocerebrosidase gene, are tightly linked to a marker in the nearby pyruvate kinase gene. This paper develops a simulation of the Ashkenazi population that considers the effects of selection and drift on the mutant allele frequency and the recombinant haplotype frequency over time. Although the fraction of mutants that are linked to the original marker decays exponentially on average, this expected value is not very likely to occur. Instead, due to random loss of the recombinant haplotype, a mutation has a significant probability of retaining complete linkage disequilibrium long after its origin, so there may be large errors in estimating the age of a mutation based on linkage data. The simulations show that the 1226G mutation probably originated between 40 and 1000 generations ago (1000 to 25,000 years ago), and the 84GG mutation probably originated between 50 and 4800 generations ago (1300 to 120,000 years ago). The recent origin of the 1226G mutation and its high current allele frequency provide strong evidence for heterozygote selection. New techniques and results developed in this paper have general applicability toward analyzing linkage disequilibrium near other mutations. For example, they potentially explain the unexpected pattern of linkage disequilibrium seen around the ΔF508 mutation of the cystic fibrosis transmembrane conductance regulator gene.

Key Words: Gaucher disease; linkage disequilibrium; heterozygote advantage; genetic drift; Ashkenazi; computer simulation.

INTRODUCTION

Disease-causing mutations are often in linkage disequilibrium with nearby polymorphic loci, usually due to their relatively recent single origin. This linkage disequilibrium is a useful tool in positional cloning and genetic diagnosis, but it also can be informative about the history of the mutation. In this study, we examine how a mutation can spread through a population and how the fraction of ancestral haplotypes might decay with time, in a finite population influenced by selection and drift (Fig. 1). We examine these factors in the context of two Gaucher disease mutations that are relatively common in the Ashkenazi Jewish population.

The general question addressed in this paper involves a population in which the allele *a* (i.e., a specific mutation) is linked to a nearby polymorphic marker at locus *B*, which contains the alleles B_1 and B_2. Initially, all copies of *a* are linked to B_1, but in every generation a fraction *r* of the alleles recombines against a background population in which B_1 has a frequency $freq(B_1)$ and B_2 has a frequency $freq(B_2)$. This continues for *g* generations after which some small fraction $x(g)$ of mutant chromosomes have a recombinant haplotype. On average, assuming that *a* is not

Correspondence and reprint requests to: F. Edward Boas, Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305. Fax: 650-725-6044. E-mail: boas@stanford.edu.
[1] Stanford University School of Medicine, Stanford, California 94305.
[2] Scripps Research Institute, La Jolla, California 92037.

1079-9796/00 $35.00

Blood Cells, Molecules, and Diseases (2000) **26**(4) August: 348–359
doi:10.1006/bcmd.2000.0314, available online at http://www.idealibrary.com on IDEAL®

$N(i)$	Breeding population size, i generations after mutation origin
a	Mutant allele at locus A
$freq(a, i)$	Proportion of chromosomes that are mutant, i generations after mutation origin
B_1	allele at locus B originally linked to a
$freq(B_1)$	Allele frequency of linked locus in the general population
B_2	unlinked allele at locus B
$freq(B_2)$	$1–freq(B_1)$
g	Age of the mutation a (generations)
r	Crossover frequency / generation, between loci A and B
$x(i)$	Proportion of mutant chromosomes that are recombinant (a linked to B_2), i generations after mutation origin
$F(p)$	Actual probability distribution for $x(g)$
$G(p)$	Probability distribution for $x(g)$ from a simulation that assumes a particular value of g
w_{Aa}	Fitness of heterozygotes for the mutation, relative to normals
w_{aa}	Fitness of homozygotes for the mutation, relative to normals

FIG. 1. Definitions of variables.

lost to fixation and $freq(B_1)$ and $freq(B_2)$ remain constant,

$$x(g) = (1 - (1 - r)^g)freq(B_2) \approx (1 - e^{-rg})freq(B_2), \quad [1]$$

regardless of population size.

Using Eq. [1], the age of the cystic fibrosis mutation in Europe (2), the age of the idiopathic torsion dystonia mutation among Ashkenazi Jews (3), and the age of the *CCR5-Δ32* AIDS-resistance allele among Caucasians (4) have been calculated. However, if there are only a small number of recombinant chromosomes in each generation, then genetic drift can make the distribution of the actual number of recombinant chromosomes highly skewed, with many populations fixed at 0% or 100% of mutant chromosomes displaying a recombinant haplotype. This means that the expected number of recombinant chromosomes may not be very representative, and a more careful analysis of the variation may be necessary.

Proposed approaches to quantitating this variation include using Luria and Delbrück's analysis of mutations in rapidly growing bacterial populations (5–7), and developing a more sophisticated mathematical treatment based on coalescent theory (8, 9). These solutions to the problem are very elegant and informative, but a simulation approach will allow for a more complex model and for analysis of more detailed statistics. The simulation developed in this paper considers the effects of both selection and genetic drift on a mutation and its decay to linkage equilibrium. Selective forces on the mutation determine how the mutation increases in frequency and thus indirectly affect the magnitude of genetic drift.

Gaucher Disease

The simulation described below will be used to examine Gaucher disease, an autosomal recessive disorder in which the lipid glucocerebroside accumulates. Manifestations can include enlargement of the spleen and liver, bone lesions, and neurological abnormalities. The disease is virtually always caused by mutations in the glucocerebrosidase (*GBA*) gene, which are uncommon in most populations, but have reached a 3.5% allele frequency among Ashkenazi Jews. The two most common *GBA* mutations in this population are 1226A→G (75%) and 84G→GG (15%). 1226G homozygotes have nonneuronopathic Gaucher disease, and their life span is typically only slightly reduced. 84GG is a null mutation, and homozygotes are non-viable (10).

We recently reported (1) that both the 1226G and 84GG mutations are in complete linkage disequilibrium with a *Pvu*II polymorphism in the *GBA* gene, suggesting that both mutations originated once in founders who were probably Ashkenazi Jews. These two mutations are also tightly linked to a polymorphic *Bsp*HI marker in the pyruvate kinase gene (*PKLR*), 71 kb distant (Table 1).

TABLE 1

Data on the Two Most Common Gaucher Mutations in the Ashkenazi Jewish Population

	GBA mutation	
	1226G	84GG
Current mutant frequency (%)	2.6	0.5
Estimated homozygote fitness	0.8	0.0
Observed haplotypes (*GBA Pvu*II site/ *PKLR Bsp*HI site)	390 −/−	55 +/+ 1 +/−
Recombinants with *PKLR*	0/390	1/56
Frequency of unlinked *PKLR* allele in general population	0.33	0.67
Recombination rate	0.00071	0.00071

Note. Homozygote fitness relative to the fitness of the normal population (w_{aa}) estimated from descriptions of clinical severity. Other data from (1). Recombination frequency assumes 1 centimorgan/megabase.

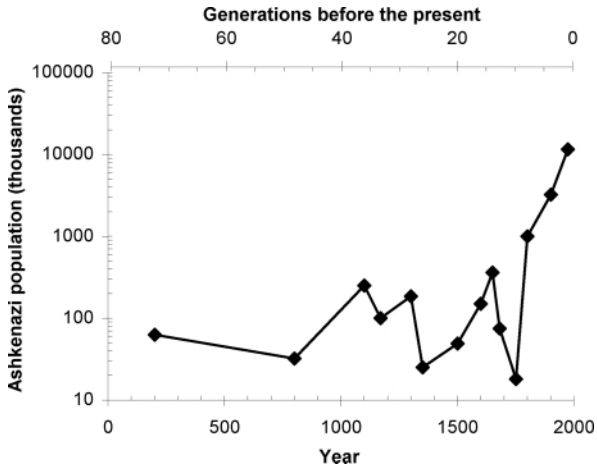

FIG. 2. Ashkenazi population size in recent history [data from (12, 13, 25, 26); conflicting estimates averaged]. This paper assumes a generation size equal to one-third the population size.

This linkage disequilibrium sets an upper limit on the ages of these two Gaucher mutations, but finding this limit requires a careful analysis of the effects of genetic drift, which will cause deviations from the expected exponential decay to linkage equilibrium. Smaller populations will experience more genetic drift, so it will be important to determine the population size as a function of time. Figure 2 shows estimates of the Ashkenazi population since the third century AD based on historical accounts. In our simulations, we assume discrete 25-year-long generations, with the generation size equal to one-third the population size. We further assume exponential growth or decay between data points, and a constant generation size outside the limits of the historical data. The latter assumption roughly agrees with the estimate of an effective generation size of 14,000 obtained from the amount of drift needed to produce the array of modern allele frequencies from putative "original" Jewish allele frequencies, after subtracting the estimated effects of migration (11).

In addition to the population data shown in Fig. 2, data regarding the variation of gene frequencies over time is needed. Some have argued that the social history of the Ashkenazi, including dramatic population expansions and contractions, makes genetic drift a major cause of the population's high frequency of genetic disorders (3, 9, 12), while others have pointed out possible selective forces such as tuberculosis (13, 14). In the case of Gaucher mutations, some undetermined heterozygote advantage provides the most likely explanation for their current high frequency, given that multiple independent mutations have reached a high frequency, and that two other lipid storage disorders caused by defective lysosomal enzymes, Tay Sachs and Niemann–Pick disease, are also relatively common in the Jewish population (15). Mutation hotspots are not a significant factor in the incidence of Gaucher alleles since many of the most common mutations are only found in a single haplotype. Our data also show that the 1226G mutation is of sufficiently recent origin that drift alone could not have elevated its frequency to current levels (see below).

MATERIALS AND METHODS

Our simulation of Gaucher disease in the Ashkenazi population uses the data described in the previous section. The general strategy is to simulate a population with selection at *GBA,* recombination between *GBA* and *PKLR,* and drift. Then, we can vary the age of the mutation and the amount of heterozygote advantage to see which values are consistent with the current observed *GBA* allele frequency and linkage disequilibrium with *PKLR.*

Mutation Frequency as a Function of Time

The simulation starts with a single mutation, g generations before the present, and first determines how the frequency of this mutation might have varied over time. Under the influence of selection, the frequency of the disease allele a changes in each generation i according to

$$freq(a,i+1) = \frac{w_{aa}freq(a,i)^2 + w_{Aa}freq(a,i) \times (1 - freq(a,i))}{w_{aa}freq(a,i)^2 + 2w_{Aa}freq(a,i) \times (1 - freq(a,i)) + (1 - freq(a,i))^2}. \quad [2]$$

We can estimate the homozygote fitness w_{aa} (relative to the normal population fitness) from clin-

Blood Cells, Molecules, and Diseases (2000) **26**(4) August: 348–359
doi:10.1006/bcmd.2000.0314, available online at http://www.idealibrary.com on IDEAL®

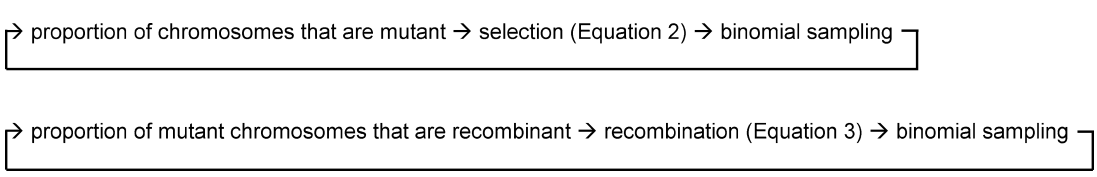

FIG. 3. Sampling schemes employed by the simulation to account for genetic drift.

ical data, but the magnitude of the heterozygote fitness w_{Aa} (also relative to the normal population fitness) is unknown. If $w_{Aa} > 1$ and $w_{Aa} > w_{aa}$, then the frequency of the disease allele increases exponentially until the appearance of homozygotes results in selection against the allele and the frequency levels off at an equilibrium value. Much of the growth of the Gaucher disease mutations is probably due to heterozygote advantage, as discussed in the previous section, but drift is still likely to play a significant role.

Thus, although the expected allele frequency in the next generation is given by Eq. [2], the actual number of mutants at generation $i + 1$ is chosen from a binomial distribution with binomial parameter $freq(a, i + 1)$ and a sample size of $2N(i)$, the number of chromosomes in the entire breeding population. Conceptually, the simulation generates a random number for each chromosome in the population to determine if it has inherited a mutant allele, but in practice, the binomial sampling is performed by a much faster algorithm from (16). This entire process is summarized in the top part of Fig. 3.

Since the actual value of w_{Aa} is unknown, the simulation randomly chooses a value of $w_{Aa} - 1$ between -5 and 5 times the value needed to produce the current mutation allele frequency in g generations in the absence of drift, thus making no assumptions about the presence or lack of heterozygote advantage. It then iterates the procedure shown in the top part of Fig. 3, and only accepts the result if the simulated allele frequency is within 5% of the actual current allele frequency. Otherwise, it repeats the process with another randomly chosen value of w_{Aa}. The distribution of accepted w_{Aa} values trails off to 0 before reaching the boundaries of the uniform distribution of proposed w_{Aa} values (Fig. 7), indicating that the distribution of proposed w_{Aa} values does not introduce any bias into the distribution of accepted w_{Aa} values. This procedure should find the most likely combination of drift and selection that is consistent with the known data on population size, w_{aa}, and current mutant frequency.

Frequency of Recombinant Chromosomes as a Function of Time

After calculating one possible trajectory for the number of mutant chromosomes as a function of time, the simulation then determines what proportion of these chromosomes have a recombinant haplotype, under the influence of recombination and drift. The number of recombinant chromosomes in each generation is chosen from a binomial distribution with an average value given by the expected number of recombinant chromosomes and a sample size given by the number of mutant chromosomes (bottom part of Fig. 3). The expected proportion of mutant chromosomes that are recombinant at generation $i+1$ is:

$$x(i + 1) = r \cdot freq(B_2) + x(i)(1 - r). \quad [3]$$

Repeating the simulation multiple times yields a probability distribution for the current fraction of mutant chromosomes that are recombinant after g generations of recombination. Specifically, this probability distribution, $G(p)$, is calculated by sorting the simulation outcomes into 100 percentile bins and assuming a uniform probability density in each bin.

Calculating the Age of the Mutation

To determine likely values for the age of the mutation, this simulation parameter is varied until there is a 95% probability that the simulated recombinant haplotype frequency is greater than (or less than) the actual measured frequency.

The probability distribution of the actual measured frequency can easily be calculated: given a

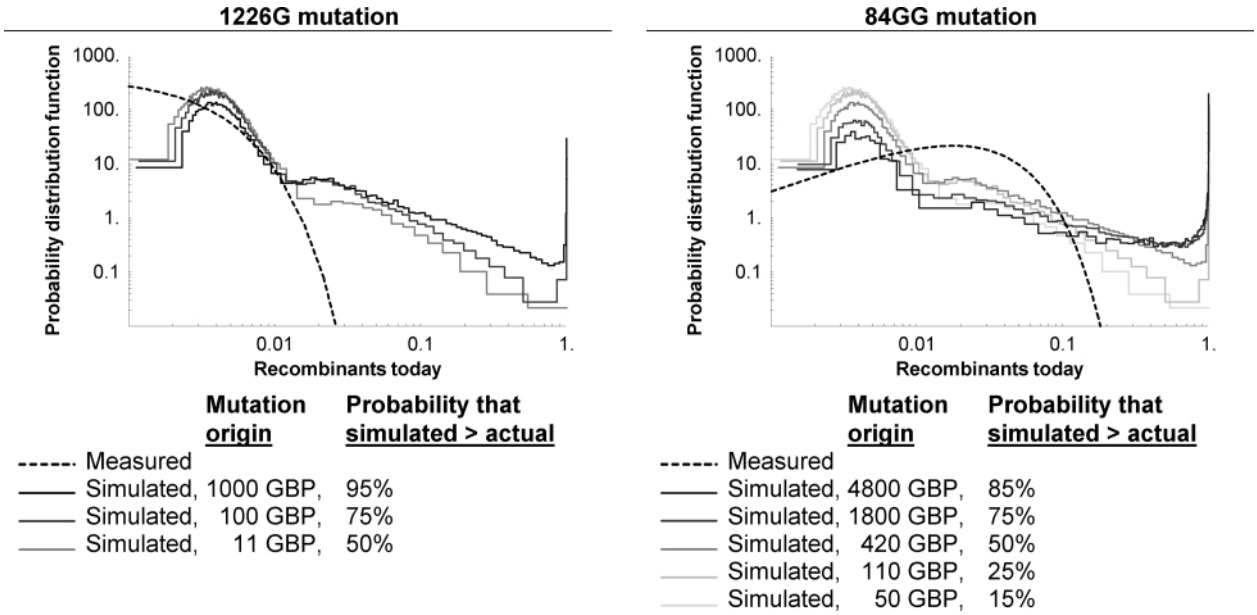

FIG. 4. Probability distributions for the current fraction of mutations in the glucocerebrosidase gene found in a recombinant haplotype with a marker in the PKLR gene. The area under each curve is 1. The dashed curve, $F(p)$, indicates the uncertainty in this fraction due to sampling error, calculated using Eq. [4] and the data in Table 1. The other curves indicate the probability distributions expected today given various ages for the mutation ($G(p)$; see figure legend). See Fig. 5 for more information. GBP, generations before the present.

uniform prior distribution for the recombinant haplotype frequency p, the posterior distribution for p, given m recombinants in n samples, is a beta distribution with parameters $\alpha = m$ and $\beta = n - m$:

$$F(p) = \frac{\binom{n}{m}p^m(1-p)^{n-m}}{\int_0^1\binom{n}{m}p^m(1-p)^{n-m}dp} = (n+1)\binom{n}{m}p^m(1-p)^{n-m}. \quad [4]$$

For large $n\hat{p}\hat{q}$, a confidence interval computed from $F(p)$ approaches the more familiar $\hat{p} \pm z \cdot \sqrt{\hat{p}\hat{q}/n}$, where $\hat{p} = m/n$, $\hat{q} = 1 - \hat{p}$, and z is the number of standard deviations from the mean needed to get the desired level of confidence in a normal distribution. Equation [4] is necessary both because the familiar approximation breaks down for the small $n\hat{p}\hat{q}$ in this problem, and because these calculations require the full probability distribution and not just a confidence interval.

Now, the probability that a recombinant haplotype frequency chosen from the simulation probability distribution $G(p)$ is greater than a frequency chosen from the actual probability distribution $F(p)$ is

$$\text{Probability}(simulated > actual) = \int_0^1\int_0^{1-x} F(p)G(p+x)dp\,dx. \quad [5]$$

This sort of explicit calculation is necessary because the distributions involved are highly nonnormal (Fig. 4) so simpler methods that assume normality will give incorrect results.

RESULTS

Figure 4 shows probability distributions for the actual recombinant haplotype frequency, and the frequency from simulations that assume various mutation ages. Notice that the frequency from the simulations often follows a bimodal distribution, with the peak at 0% corresponding to the

Blood Cells, Molecules, and Diseases (2000) **26**(4) August: 348–359
doi:10.1006/bcmd.2000.0314, available online at http://www.idealibrary.com on IDEAL®

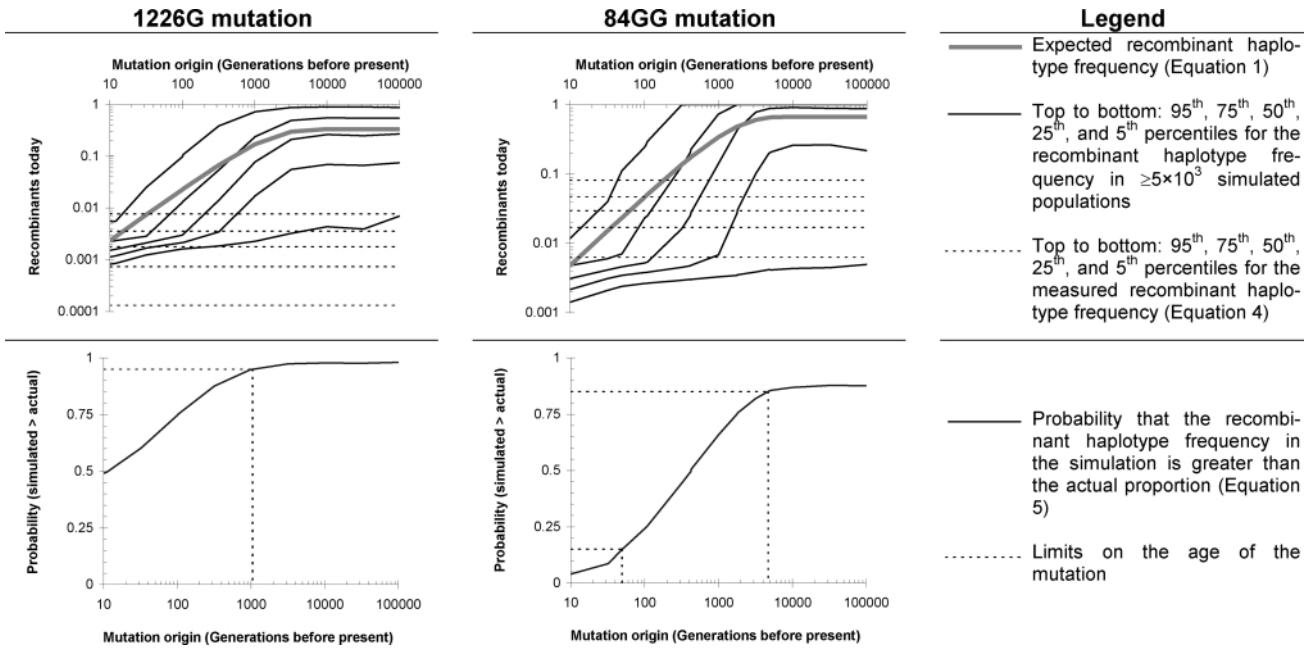

FIG. 5. Expected current fraction of mutant chromosomes that are recombinant, as a function of the mutation age. The top panel compares the current measured recombinant haplotype frequency with the recombinant haplotype frequencies from simulations with varying dates of mutation origin. The bottom panel calculates the probability that the simulation indicates a greater recombinant haplotype frequency than is actually the case. These probabilities, calculated with Eq. [5], indicate that the 1226G mutation most likely originated less than 1000 generations before the present (95% confidence; dotted lines, bottom left graph) and the 84GG mutation most likely originated between 50 and 4800 generations before the present (85% confidence; dotted lines, bottom right graph).

ancestral haplotype, and the peak at 100% corresponding to the recombinant haplotype. This suggests that our careful work to elucidate the full distribution of recombinants expected, rather than just the mean and variance, has been necessary. As expected, older mutations have a smaller probability of falling in the peak at 0%, and a greater probability of falling in the peak at 100%.

The top panel of Fig. 5 displays the percentiles for the probability distributions from Fig. 4. Importantly, the percentiles demonstrate the extreme amount of random variation possible in the speed of decay to linkage equilibrium. For older mutations, the percentiles for the expected fraction of mutant chromosomes that are recombinant sweep above the percentiles for the actual fraction, making them less consistent with each other. The probability that the simulated fraction is greater than the actual fraction, graphed in the bottom panel of Fig. 5, indicates that at 95% confidence, the 1226G mutation is younger than 1000 generations. The probabilities for the 84GG mutation never reach 95%, because of the large variation in the frequency of mutant chromosomes that are recombinant, even after the expected frequency has reached equilibrium. Therefore, we must settle for a lower confidence level: at 85% confidence, the 84GG mutation is between 50 and 4800 generations old.

Figure 6 shows the details of the simulations using the maximum likely value for g for the two mutations. The top row of Fig. 6 shows how the mutation allele frequency might have increased to its current value from $1/(2N(0))$. Note that the mutant frequency in most of the simulated populations lies above the expected curve for a population without drift, indicating that upward drift is greatest immediately after the mutation's origin. Populations whose mutation allele frequency happens to drift upward at the beginning are more likely to avoid immediate fixation at 0, and are hence more likely to reach the current mutant frequency. 84GG probably decreased in frequency recently because a mutation with a higher sustained frequency would be more likely to have survived to the present.

doi:10.1006/bcmd.2000.0314, available online at http://www.idealibrary.com on IDEAL®

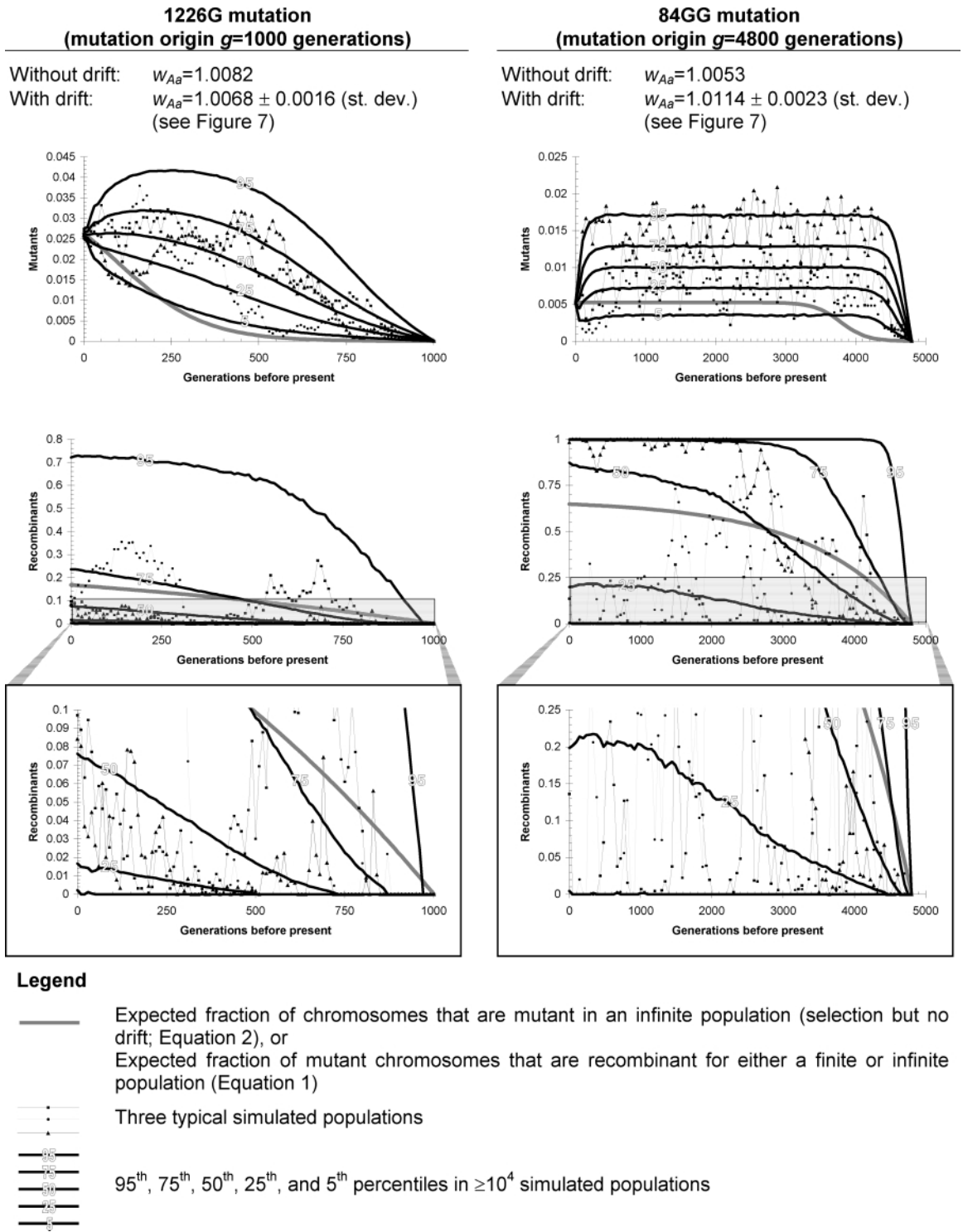

FIG. 6. Simulations of the mutant allele frequency and the recombinant haplotype frequency as a function of time, assuming the mutations originated at their maximum likely age (calculated in Fig. 5).

After accounting for drift in the finite Ashkenazi population, an average heterozygote advantage of 0.68% is needed to raise 1226G to its current frequency in 1000 generations, and an average heterozygote advantage of 1.1% is needed to raise 84GG to its current frequency in 4800 generations (Fig. 7). If the mutations had originated earlier than their maximum likely age,

Blood Cells, Molecules, and Diseases (2000) **26**(4) August: 348–359
doi:10.1006/bcmd.2000.0314, available online at http://www.idealibrary.com on IDEAL®

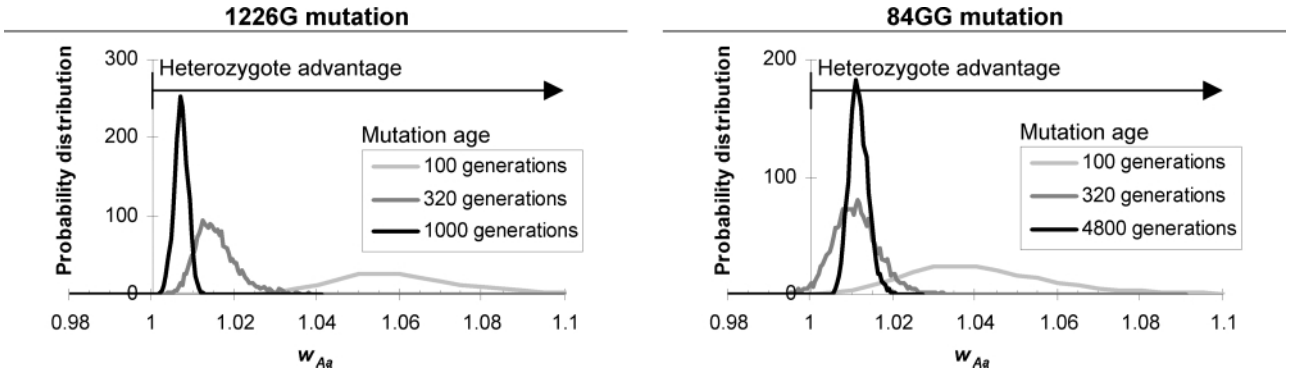

FIG. 7. Distribution of values for the heterozygote fitness w_{Aa} obtained by assuming various values for the mutation age within the range calculated earlier. Values for w_{Aa} are proposed from wider uniform distributions, and only the values which allow the simulation to reach the current allele frequency are accepted. Simulations of the 1226G mutation never had $w_{Aa} \leq 1$, out of $\geq 5 \times 10^3$ trials at each of several different mutation ages. Simulations of the 84GG mutation had $w_{Aa} \leq 1$ a maximum of 2.1% of the time, for the simulation of a 320-generation-old mutation.

even greater heterozygote selection would be necessary. Furthermore, 1226G/84GG heterozygotes have a low fitness, perhaps around 0.25. After correcting for this (and assuming modern allele frequencies), 1226G/normal heterozygotes have an expected fitness of at least 1.1% above normal, and 84GG/normal heterozygotes have an expected fitness of at least 3.2% above normal. At the end of this section we will present evidence that drift alone could not have elevated 1226G to its current allele frequency, supporting that idea that this heterozygote advantage is real.

The bottom two rows of Fig. 6 show, on two different scales, the decay in linkage disequilibrium between the mutation and the marker over time. On average, the linkage disequilibrium decays exponentially, but for the small mutant population present soon after its origin, this expected value is achieved by a significant probability that no recombinant chromosomes will occur, balanced by a small probability that a fortuitous recombination event will create a large fraction of recombinant chromosomes. Thus, the 95th percentile of the fraction of recombinant chromosomes starts increasing rapidly soon after the mutation origin, while the 25th and 5th percentiles remain at 0 until the population is large enough to decrease the genetic drift that pushes allele frequencies there. This observation emphasizes the need to examine population data and possible scenarios for the growth of the mutation allele frequency, as we have done, because the number of mutant chromosomes influences the variation in the rate at which linkage disequilibrium decays.

The preceding analysis cannot be used to find a lower limit for the age of the 1226G mutation, because no recombinants have ever been observed with this mutation (1), which would be consistent with an arbitrarily recent origin.

Instead, we can calculate a minimum bound for the age of 1226G by determining the least amount of time needed for it to reach its current allele frequency under ideal conditions. Sickle cell anemia, the best-documented example of heterozygote advantage, reaches $w_{Aa} = 1.25$ at the highest levels of malaria (17). If we set this as the maximum reasonable heterozygote advantage for Gaucher mutations, the mutation would require around 40 generations to reach its current frequency.

Given these limits on the ages of the Gaucher mutations, we can test to see if they could have reached their current frequencies in the Ashkenazi population by drift alone. For the sake of argument, we assume that there is selection against homozygotes (Table 1) but no selection for or against heterozygotes. This is a conservative assumption since there is in fact strong selection against 1226G/84GG heterozygotes. Given these assumptions, we iterated the procedure shown in the top part of Fig. 3 to determine the fate of 1226G or 84GG mutations as a function of how

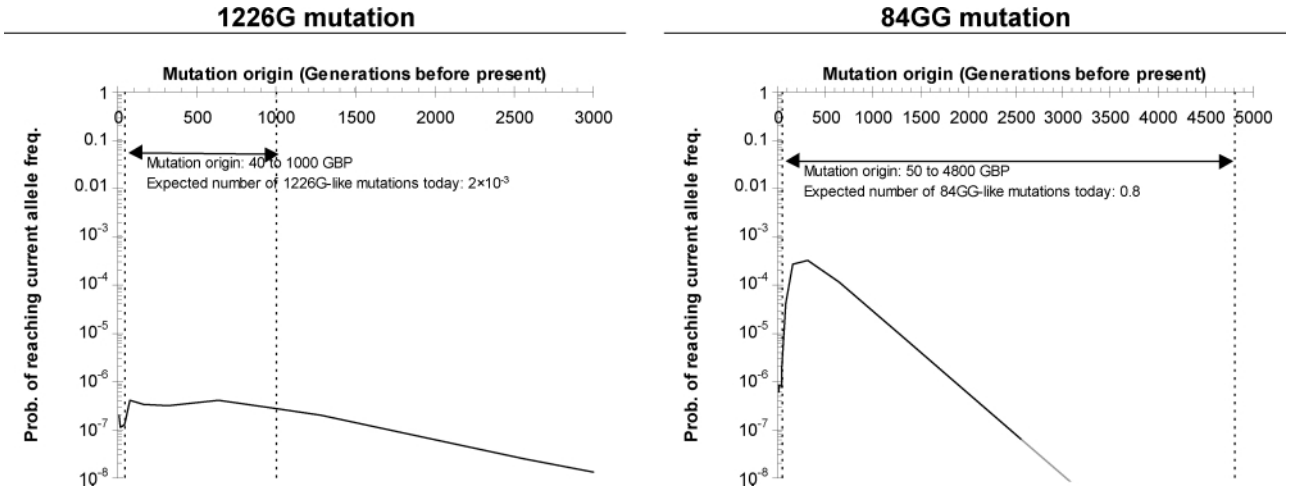

FIG. 8. Probability that a 1226G or 84GG mutation reaches or exceeds its current allele frequency, assuming that the mutations do not confer a heterozygote advantage. The gray portion of the 84GG mutation graph represents extrapolated data. The very low probabilities for the 1226G mutation reaching its current allele frequency indicate that this assumption is almost certainly invalid for 1226G.

long ago they originated (Fig. 8). Both graphs have a single peak in probability. The probability of reaching the current allele frequency decreases for mutations of recent origin because there has not been enough time for the mutation to drift upwards in frequency. The probability of reaching the current allele frequency also decreases for mutations of ancient origin because of fixation at zero frequency.

The data in Fig. 8 provide strong evidence of heterozygote selection for the 1226G mutation in the Ashkenazi population. A typical mutation rate of 0.33×10^{-8}/year/nucleotide (18) translates into a mutation rate of 10^{-4}/generation in the glucocerebrosidase gene. Multiplying this rate by the number of alleles (calculated from the Ashkenazi population estimates in Fig. 2) and the probabilities from Fig. 8, then integrating over the estimated range of mutation ages (dotted lines in Fig. 8), will approximate the expected number of *any* glucocerebrosidase mutation (not just 1226G or 84GG) being present today at 1226G or 84GG allele frequencies. The result is that we would expect 2×10^{-3} 1226G-like mutations today and 0.8 84GG-like mutations today in the absence of heterozygote selection. Even without using the calculated age limits, we can extrapolate and integrate to infinity, yielding 3×10^{-3} 1226G-like mutations today that originated at *any* time. Thus, our original assumption must be wrong: it would be very unlikely for the 1226G mutation, or any other mutation like it, to have reached its current allele frequency without any heterozygote advantage.

DISCUSSION

We have shown that the linkage disequilibrium between a mutation and a nearby marker can serve as a molecular clock for the age of the mutation, but only after accounting for some of the clock's peculiar properties. For example, this "clock" can easily get stuck at zero in small populations or closely linked genes due to random loss of recombinant chromosomes. This paper's analysis of the variations in the clock speed allows us to better interpret linkage data.

An interesting corollary of this analysis is that markers extremely close to a locus of interest may stay more tightly linked to it than the expected exponential decay of linkage disequilibrium with distance. This phenomenon has been observed with markers near the ΔF508 mutation of the cystic fibrosis transmembrane conductance regulator gene (19; Fig. 9).

Some authors have concluded that the recent extreme fluctuations in Ashkenazi population size have created a series of founder effects that might

Blood Cells, Molecules, and Diseases (2000) **26**(4) August: 348–359
doi:10.1006/bcmd.2000.0314, available online at http://www.idealibrary.com on IDEAL®

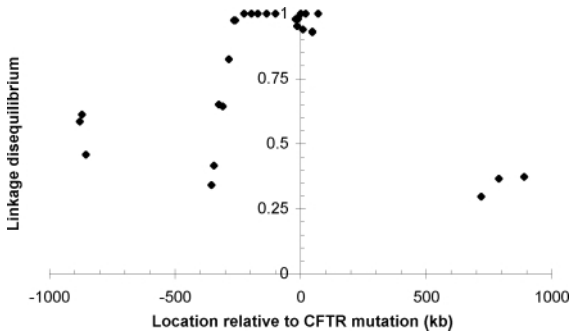

FIG. 9. Linkage disequilibrium near the ΔF508 mutation of the CFTR gene. Data compiled in (19).

have allowed rare mutations to become more common in this population compared to many others (3, 12). However, our simulations (top row of Fig. 6) indicate that for the 1226G and 84GG mutations in the Ashkenazi population, mutant frequencies do not change faster during population bottlenecks. In the context of the potentially long histories of the mutations, these recent bottlenecks reduced the population size to levels that had existed for hundreds and maybe even thousands of years. The change in allele frequencies during these population crashes were thus no larger than genetic drift in the small Ashkenazi populations of several hundred years ago.

Notably, the linkage data on the 1226G mutation indicates that it is too young to have reached its current frequency by drift alone; some form of heterozygote advantage must be operating.

Some recent studies (20–22) have calculated that the 1226G mutation originated 25–280 generations ago, and the 84GG mutation originated 56 generations ago. While these numbers are roughly consistent with the estimates presented in this paper, they are near the lower bounds. Our simulations indicate (Fig. 5) that these mutations could in fact be substantially older than a simple calculation might suggest.

Our model, as with any model of a human population, oversimplifies the situation. Several complicating factors are worth pointing out.

1. Single vs fragmented population. Our simulation assumes a single population, but in reality, the Ashkenazi population was probably fragmented. If the migration rate between subpopulations is low compared to the recombination rate, it will retard the decay to linkage equilibrium (23). Indeed, it is possible for subpopulations to be in complete linkage equilibrium while the population at a whole is not at equilibrium. However, the low recombination rate between the *GBA* and *PKLR* genes (0.071%) means that even low levels of migration between subpopulations would be sufficient to make Ashkenazi Jews behave like a single population with respect to linkage disequilibrium between these two loci.

2. Discrete vs continuous generations. Our assumption of discrete generations should not significantly affect our results since the time scale of the simulation is much longer than the generation length. If anything, the assumption of discrete generations would tend to overestimate the possible variation [analogous to a result in (6)].

3. Nonuniform population expansion. We assume that the transfer of genes to the next generation can be modeled by a random sampling process, meaning that the number of copies of a gene in the next generation is approximately Poisson distributed with a mean and variance equal to the population growth rate per generation. It is possible that in the real Ashkenazi population, the variation in the number of copies of a gene passed on to the next generation is greater than this. For example, higher Jewish social classes have historically had more children than lower social classes. This would increase the magnitude of genetic drift.

4. Constant vs fluctuating selection. If the heterozygote advantage of *GBA* mutations depends on sporadic events such as tuberculosis outbreaks or famines, then the amount of selection might fluctuate from generation to generation.

5. Drift in the linked allele frequency. This would increase the uncertainty in the rate of observed recombination.

6. Inbreeding. This would increase selection against Gaucher mutations by increasing the incidence of homozygotes. However, the inbreeding coefficient for Ashkenazi Jews is small, around 0.8% (11).

7. Migration. There is mixed evidence on the amount of migration into the Ashkenazi popula-

tion. One study suggests a relatively large (relative to other Jewish populations) "cumulative" migration rate of 54% based on the degree of similarity in allele frequencies between the Ashkenazi and neighboring populations (11), while another suggests little admixture over the past 700 years (24). The existence of a well-defined Ashkenazi population in the distant past is even less certain.

ACKNOWLEDGMENTS

I thank James Koziol and Ernest Beutler for many productive discussions and suggestions on the manuscript. My deepest gratitude goes to Dr. Beutler for the invaluable guidance he has provided me as my mentor over the past few years.

REFERENCES

1. Demina, A., Boas, E., and Beutler, E. (1998) Structure and linkage relationships of the region containing the human L-type pyruvate kinase (PKLR) and glucocerebrosidase (GBA) genes. *Hematopathol. Mol. Hematol.* **11**(2), 63–71.
2. Serre, J. L., Simon-Bouy, B., Mornet, E., Jaume-Roig, B., Balassopoulou, A., Schwartz, M., Taillandier, A., Boue, J., and Boue, A. (1990) Studies of RFLP closely linked to the cystic fibrosis locus throughout Europe lead to new considerations in population genetics. *Hum. Genet.* **84,** 449–454.
3. Risch, N., de Leon, D., Ozelius, L., Kramer, P., Almasy, L., Singer, B., Fahn, S., Breakefield, X., and Bressman, S. (1995) Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nat. Genet.* **9,** 152–159.
4. Stephens, J. C., Reich, D. E., Goldstein, D. B., Shin, H. D., Smith, M. W., Dean, M., *et al.* (1998) Dating the origin of the *CCR5-Δ32* AIDS-resistance allele by the coalescence of haplotypes. *Am. J. Hum. Genet.* **62,** 1507–1515.
5. Hästabacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A., and Lander, E. (1992) Linkage disequilibrium mapping in isolated founder populations: Diastrophic dysplasia in Finland. *Nat. Genet.* **2,** 204–211.
6. Sarkar, S. (1991) Haldane's solution of the Luria–Delbrück distribution. *Genetics* **127,** 257–261.
7. Luria, S. E., and Delbrück, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28,** 491–511.
8. Rannala, B., and Slatkin, M. (1998) Likelihood analysis of disequilibrium mapping, and related problems. *Am. J. Hum. Genet.* **62,** 459–473.
9. Thompson, E. A., and Neel, J. V. (1997) Allelic disequilibrium and allele frequency distribution as a function of social and demographic history. *Am. J. Hum. Genet.* **60,** 197–204.
10. Beutler, E., and Gelbart, T. (1997) Hematologically important mutations: Gaucher disease. *Blood Cells Mol. Dis.* **23,** 2–7.
11. Carmelli, D., and Cavalli-Sforza, L. L. (1979) The genetic origin of the Jews: A multivariate approach. *Hum. Biol.* **51**(1), 41–61.
12. Fraikor, A. L. (1977) Tay–Sachs disease: Genetic drift among the Ashkenazim Jews. *Soc. Biol.* **24,** 117–134.
13. Motulsky, A. G. (1979) Possible selective effects of urbanization of Ashkenazi Jews. *In* Genetic Diseases among Ashkenazi Jews (Goodman, R. M., and Motulsky, A. G., Eds.), pp. 301–312. Raven Press, New York.
14. Rotter, J. I., and Diamond, J. M. (1987) What maintains the frequencies of human genetic diseases? *Nature* **329,** 289–290.
15. Diamond, J. M. (1994) Jewish lysosomes. *Nature* **368,** 291–292.
16. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992) Numerical Recipes in C: The Art of Scientific Computing, 2nd ed. Cambridge UP, Cambridge.
17. Wiesenfeld, S. L. (1967) Sickle-cell trait in human biological and cultural evolution. *Science* **157,** 1134–1140.
18. Ohta, T., and Kimura, M. (1971) Functional organization of genetic material as a product of molecular evolution. *Nature* **233,** 118–119.
19. Collins, A., and Morton, N. E. (1998) Mapping a disease locus by allelic association. *Proc. Natl. Acad. Sci. USA* **95,** 1741–1745.
20. Diaz, G. A., Gelb, B. D., Risch, N., Nygaard, T. G., Frisch, A., Desnick, R. J., *et al.* (2000) Gaucher disease: The origins of the Ashkenazi Jewish N370S and 84GG acid β-glucosidase mutations. *Am. J. Hum. Genet.* **66,** 1821–1832.
21. Colombo, R. (2000) Age estimate of the N370S mutation causing Gaucher disease in Ashkenazi Jews and European populations: A reappraisal of haplotype data. *Am. J. Hum. Genet.* **66,** 692–697.
22. Díaz, A., Montfort, M., Cormand, B., Zeng, B., Pastores, G. M., Grinberg, D., *et al.* (1999) Gaucher disease: The N370S mutation in Ashkenazi Jewish and Spanish patients has a common origin and arose several thousand years ago. *Am. J. Hum. Genet.* **64,** 1233–1238.
23. Nei, M., and Li, W. H. (1973) Linkage disequilibrium in subdivided populations. *Genetics* **75,** 213–219.

Blood Cells, Molecules, and Diseases (2000) **26**(4) August: 348–359
doi:10.1006/bcmd.2000.0314, available online at http://www.idealibrary.com on IDEAL®

24. Bonné-Tamir, B., Ashbel, S., and Kenett, R. (1979) Genetic markers: Benign and normal traits of Ashkenazi Jews. *In* Genetic Diseases among Ashkenazi Jews (Goodman, R. M., and Motulsky, A. G., Eds.), pp. 59–76. Raven Press, New York.
25. Ankori, Z. (1979) Origins and history of Ashkenazi Jewry (8th to 18th century). *In* Genetic Diseases among Ashkenazi Jews (Goodman, R. M., and Motulsky, A. G., Eds.), pp. 19–46. Raven Press, New York.
26. Neel, J. V. (1979) History and the Tay Sachs allele. *In* Genetic Diseases among Ashkenazi Jews (Goodman, R. M., and Motulsky, A. G., Eds.), pp. 285–99. Raven Press, New York.