Proc. Natl. Acad. Sci. USA
Vol. 75, No. 10, pp. 5585–5588, November 1978
Genetics

Indirect estimates of mutation rates in tribal Amerindians

(electrophoretic variants/tribal polymorphisms/random genetic loss/mutation pressure)

JAMES V. NEEL* AND EDWARD D. ROTHMAN*†

* Department of Human Genetics and † Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109

Contributed by James V. Neel, July 10, 1978

ABSTRACT **Three different formulations have been used to estimate the average rate/locus/generation with which mutation results in electrophoretically detectable variants of 28 proteins in 12 tribal Amerindian populations. All methods are indirect—i.e., they assume a reasonable approximation to equilibrium between mutation and loss of mutants from the population—and are based on the further assumption that the biochemical traits under consideration are essentially neutral in their phenotypic effects. Despite the fact that the methods draw on somewhat different aspects of the available data, there is satisfactory agreement between them, the average of the three estimates being 1.6×10^{-5}/locus/generation. This average does not encompass mutation that does not result in a change in electrophoretic mobility or that results in loss of enzyme activity. It is noteworthy that this estimate is in satisfactory agreement with a recent estimate by Neel and Thompson [Neel, J. V. & Thompson, E. A. (1978) *Proc. Natl. Acad. Sci. USA* 75, 1904–1908] of the mutation pressure necessary to maintain the number of "private" genetic polymorphisms being encountered in Amerindian tribes, if selection is not a factor.**

The rate at which genes mutate is one of the fundamental parameters of genetics. Until recently, the techniques for measuring these rates in higher organisms have been keyed to rather gross phenotypic events, such as death of the developing embryo and obvious departures from normal appearance. With the development of convenient and inexpensive ways of detecting variant proteins, the possibility now exists of carrying the study of mutation rates to the protein level. Several years ago, data on protein variants in six relatively intact and unmixed Amerindian tribes were utilized to generate a first indirect estimate for man of the rate at which mutation resulted in such variants (1). That estimate, to be discussed in some detail later, was approximately an order of magnitude higher than the rates adduced by some students of human genetics from morphological data (2–4). Since that time, several new approaches to the indirect estimation of mutation rates from data of this type have been proposed (5, 6) and the data base has been approximately doubled. This communication will take advantage of these recent developments to revise and extend the earlier estimate.

THE DATA

During the past 14 years, we have used the technique of electrophoresis to survey 12 Amerindian tribes for the occurrence of variants of various proteins of the blood serum and erythrocytes, the exact number of proteins investigated varying from 13 for the Xavante, the first tribe to be studied, to 28 for the more recently studied tribes. The findings have recently been reviewed in detail (7). Table 1 presents the data concerning the tribal samples that are necessary for an indirect calculation of mutation rates by the approaches to be used in this communication. A critical parameter is the number in the adult generation in each tribe (N). Studies on one of these 12 tribes, the Yanomama, have demonstrated that 48% of the population fall into the age interval 15–40 yr, which we equate to adulthood (8). The number of adults in each tribe has accordingly been calculated by applying that percentage to the estimated tribal size. From the total number of electrophoretic determinations for each tribe, we can calculate the average number of persons sampled per protein, the number of adults sampled, and the percentage of N that this number represents.

Table 2, extracted from tables 1 and 3 of ref. 7, classifies the variants encountered in these tribes. Four conventions have been followed in the entries. (*i*) The "widely distributed polymorphisms" include not only the five present in virtually all the tribes (involving haptoglobin, erythrocyte acid phosphatase, phosphoglucomutase 1, esterase D, and galactose-1-phosphate uridyltransferase) but also the less commonly encountered transferrin D_{Chi} and the 6-phosphogluconate dehydrogenase C variants. (*ii*) When an electrophoretically identical variant is found in multiple tribes, it is credited to the tribe of highest frequency and appears in the table only once. Because of the extension of material since the previous calculation of mutation rates, this procedure—which we regard as a conservative step—results in the tribal reassignment of some variants; for this reason the number of unique variants by tribe in this publication will not always agree with that in the previous estimate of Amerindian mutation rates (1). (*iii*) Because the numbers that enter into the computation are based upon an estimate of the "generation of reproducing adults," any variant in table 1 of ref. 7 that was not present in at least one adult has not been included in Table 2 of this paper. (*iv*) An unusual feature of these studies has been the frequency of occurrence of "private polymorphisms"—i.e., alleles with frequencies equal to or greater than 0.01 completely or largely confined (as far as is now known) to a single tribe. These "private polymorphisms," which have recently been the basis for a treatment of the probability of a founder effect and the mutation rate implied by such polymorphisms, have been listed for ease of cross-reference to data presented elsewhere (9, 10).

ESTIMATION OF TOTAL NUMBER OF VARIANTS IN A POPULATION

In all the studies, the unit of tribal sampling has been the village, within which the blood samples have been obtained from as many persons as possible. Because for all tribes only a minority of the villages has been sampled, and hence only a minority of the population, it is clear that, although most of the variants with the frequency of private polymorphisms should have been detected in these tribes, many variants present only in one or several families were undoubtedly missed. Rothman and Adams (6), in the paper preceding this contribution, have presented a statistical approach to estimating the average number of variants per locus in a tribe (K) from the observed number per locus (k), the sample size (n), and N as previously defined. Their approach is based on the following computation:

$$E(k|K,2N) = K\left[1 - \frac{\Sigma g(i)\binom{2N-i}{2n}}{\binom{2N}{2n}}\right] \qquad [1]$$

in which $E(k|K,2N)$ is the conditional expectation of k, given

Table 1. Percentage samples from each of 12 tribes used as basis for calculation of mutation rates

Tribe	Estimated total population	Approximate N	Total polypeptides examined, no.	Total system determinations, no.	Av. no. sampled/polypeptide	Estimated adults sampled, no.	% N sampled
Ayoreo	1,500	720	27	5,449	202	97	13.47
Baniwa	1,500	720	27	10,177	377	181	25.14
Cayapo	1,500	720	28	15,253	545	262	36.39
Guaymí	30,000	14,400	27	13,100	485	233	1.62
Kraho	600	288	28	5,348	191	92	31.94
Macushi	4,000	1,920	27	13,476	499	240	12.50
Makiritare	1,500	720	27	13,926	516	248	34.44
Panoa	18,000	8,640	27	9,025	334	160	1.85
Piaroa	3,000	1,440	24	3,502	146	70	4.86
Wapishana	2,000	960	28	17,221	615	295	30.73
Xavante	1,700	816	13	4,218	324	156	19.12
Yanomama	15,000	7,200	28	55,968	1,999	960	13.33

both K and $2N$, the symbol $\binom{a}{b}$ represents $[a!/b!(a-b)!]$, and $g(i)$ represent the proportions of alleles that occur exactly i times in the population.

A model, the geometric series (see ref. 11), describing the number of copies of a mutant allele transmitted to the next generation, is fitted to observed data and then, in conjunction with assumptions given in ref. 6, is used to obtain estimates of $g(i)$. Then, replacing the conditional expectation in the above formula with the observed number of alleles and substituting our estimates, $\tilde{g}(i)$, an extrapolation rule is obtained. It is

$$\hat{K} = \frac{k}{1 - \sum \tilde{g}(i)\binom{2N-i}{2n} \Big/ \binom{2N}{2n}}. \quad [2]$$

Because $g(i)$ is a rapidly decreasing sequence in i and the sampling fraction, $f = n/N$, is usually small, this computation is well approximated in our applications by

$$\hat{K} = \frac{k}{1 - \sum_{i=1}^{30} \tilde{g}(i)(1-f)^i}. \quad [3]$$

Columns 5 and 6 of Table 2 present, by tribe, values for k and K. These values are averages, the number of polypeptides involved in the estimate ranging from 13 to 28 (Table 1).

Table 2. Summary of kinds of variants detected by electrophoresis in 12 Amerindian tribes under consideration*

Tribes	Widely distributed polymorphisms	Private polymorphisms	Rare variants	Averages k	Averages K
Ayoreo	3	0	0	0	0
Baniwa	6	1	1	0.0741	0.1415
Cayapo	5	1	1	0.0714	0.1126
Guaymí	7	1	1	0.0741	1.1475
Kraho	4	0	1	0.0357	0.0601
Macushi	5	1	2	0.1111	0.3230
Makiritare	5	0	2	0.0741	0.1200
Panoa	5	1	0	0.0370	0.5074
Piaroa	6	0	1	0.0417	0.2448
Wapishana	6	2	3	0.1786	0.3066
Xavante	2†	0	2	0.1538	0.3430
Yanomama	5	1	0	0.0357	0.0995

* Extracted from ref. 8.
† No determination of ESD, GALT, or PGM_1.

All three of the procedures for estimating mutation rates to be utilized require the following assumptions:

(*i*) The tribe is the breeding unit. Although, as has been documented for the tribes under discussion (12–14), there is no doubt of the occasional capture and incorporation into a tribe of members of other tribes or the absorption of the remnants of a decimated tribe, the apparent restriction of so many private polymorphisms to single tribes suggests that, in the main, tribal boundaries have been effective barriers to migration. Our convention of introducing each variant into the calculation only once, according to the tribe of its highest frequency, will obviate double counting for a variant carried from one tribe to another, at least for the tribes surveyed. This is a conservative procedure because it assumes that identical electromorphs in separated tribes stem from the same mutational event, an assumption not necessarily true.

(*ii*) The tribe has been the breeding unit for a sufficient period of time that there is an approximation to genetic equilibrium. The evidence for this again rests largely on the restriction of the private polymorphisms to single tribes or to adjacent tribes when recent admixture is certain or probable. However, as pointed out elsewhere (15), genetic equilibrium in the classical sense has probably never been attained in the demes of these tribal populations (or in any other human population).

(*iii*) Tribal numbers have been relatively constant since the tribe was established. This is clearly incorrect for most tribes. The founding populations were undoubtedly smaller than the present populations. Furthermore, there is good evidence that two of these tribes (Guaymí and Yanomama) have recently expanded rather considerably (13, 16) and it is our impression that others of these tribes (Baniwa, Kraho, Makiritare, and Pano) have been somewhat decimated in consequence of their contacts with neo-Americans. Unfortunately, for none of these tribes are there any *precise* data on growth or numerical fluctuations during the remote or even recent past; for these purposes we will assume that present numbers can be equated to the mean throughout the tribal history. Thompson and Neel (10) in a recent treatment concluded that the effect of cyclic variation in population size generally was to decrease the probability of survival of an individual mutant but to increase the numbers with which those surviving were represented. Because all three of the procedures for estimating mutation rates to be discussed depend on the number of different mutations represented in the population and because some cyclic variation in population size is almost certain, cyclic variation should render the present estimates somewhat conservative.

The three procedures that we have used for the estimation of mutation rates are as follows.

1. **The Method of Kimura and Ohta.** The rate of mutation (17) per locus per generation (μ) is given by

$$\mu = \frac{K}{2N} \cdot \frac{1}{\bar{t}_0} \quad [4]$$

in which K and N are as previously defined and $\bar{t}_0$ is the average mutant survival time in generations for those mutants not going to fixation. Li *et al.* (11) have recently described a Monte Carlo simulation of an Indian population. Among other estimates, for introduced mutations assumed destined for ultimate loss, this simulation yielded an estimate of the mean number of copies of a mutant gene left by a mutant carrier, m, of 0.978. Because, in Table 2, variants have been enumerated in the adult generation, the value of $\bar{t}_0$ necessary in calculations using Eq. 4 is for newly arisen mutations whose bearer survives to adulthood. In the tribe studied in greatest detail, the Yanomama, the value of c, the geometric parameter in the equation describing the distribution of mutant-bearing offspring surviving to adulthood, was 0.40 (18). From these values of m and c, we calculate $\bar{t}_{0_A} = 5.6$ generations.

Although Kimura and Ohta (17) defined K to include all the mutant alleles present in the population, Neel (1) and Nei (5) have expressed reservations concerning the inclusion, in a calculation involving tribal groups, of the alleles defining polymorphisms of wide distribution (in this case, Pan-Amerindian), and we shall exclude them from the calculation. This is another conservative aspect of this calculation. To apply this format we must correct k as described in the previous section.

2. **The Method of Nei.** The basic relationship (5) of interest in the present context is:

$$k = 4N \log_e(2nq) \quad [5]$$

in which N, n, k, and μ are as defined above and q is the average gene frequency across populations above which an allele will be considered a polymorphism of wide distribution, possibly or probably maintained by selection and so to be excluded from the calculation. Nei assigned q a value of 0.01. To illustrate the parallelism between this approach and Eq. 4, we rewrite Eq. 5 as

$$\mu = \frac{k}{2N} \cdot \frac{1}{2 \log_e(2nq)}. \quad [6]$$

If the total population were sampled, as assumed in Eq. 4, then $k = K$ as defined above and $n = N$, and Eq. 5 becomes

$$\mu = \frac{K}{2N} \cdot \frac{1}{2 \log_e(2Nq)}. \quad [7]$$

In this form, $2 \log_e(2Nq)$ is the equivalent of $\bar{t}_0$ in Eq. 4. For a population in which the mean number of offspring is 2, the variance in offspring number is 2, and there is no death up to and through the reproductive period, Kimura and Ohta (17) find

$$\bar{t}_0 = 2 \frac{N_e}{N} \log_e(2N), \quad [8]$$

in which N_e = "effective"population size. We (11) have demonstrated both analytically and through simulation that the population structure of the Yanomama results in an expected time to extinction ($\bar{t}_0$) much shorter than that predicted by Eq. 8. The Nei formulation apparently accepts Eq. 8 and equates $N_e = N$ but then introduces an arbitrary correction, $2 \log_e q$, into the equivalent expression for $\bar{t}_0$, to "compensate" for the exclusion from the calculation of variant alleles with frequencies greater than 0.01.

3. **The Method of Rothman and Adams.** A general model connecting the expected number of alleles with the frequency i [$G(i)$], the mutation rate (μ), and the conditional probability of transmitting j copies of a mutant allele from i copies of this allele (P_{ij}) may be deduced from the assumptions given in ref. 6. It is

$$2N\mu + \sum_{i>1} G(i)P_{i1} = G(1)[1 - P_{i1}]. \quad [9]$$

The equation represents the balance, at equilibrium, between the expected number of alleles entering the singleton class and those alleles that exit. This latter quantity is presented on the right-hand side whereas the former is composed of two portions: the first term, $2N\mu$, represents the expected number of mutants produced in a single generation, which, in view of our infinite alleles structure, appear as single copies; and the second term represents the drift from alleles at higher frequency into the singleton class.

Rearranging the terms in this expression and replacing the unknown $G(i)$ and P_{i1}, $i = 1, 2, \ldots$, with estimates described below, an estimator for the mutation rate follows:

$$\hat{\mu} = \frac{\hat{K}}{2N} [\tilde{g}(1) - \tilde{\Sigma} g(i)P_{i1}]. \quad [10]$$

The sequences $\tilde{g}(i)$, $[g(i)K = G(i)]$, and P_{i1}, i = 1, 2, . . . , $2N$, are estimated from the number of copies transmitted by a single mutant allele and then $\hat{K}$ is obtained from our extrapolation formula.

The results of applying these procedures to the data are presented in Table 3. To prevent the data from a single extensively studied tribe, such as the Yanomama, from dominating the results, all averages are unweighted. Considered tribe by tribe, the results yielded by Eqs. **4** and **10** are more similar to each other than those yielded by Eq. **5** are to either. However, the means of the estimates yielded by the three approaches are remarkably similar, given the different aspects of the data utilized by each of the approaches. Furthermore, given the large variations in both numbers and percentage of tribe sampled, we are impressed by the relative consistency in the results. We will work with the average of all three estimates, namely, 1.6×10^{-5}/locus per generation.

Table 3. Results of applying three different procedures for indirect estimation of mutation rates

Tribes	Kimura–Ohta formulation ($\bar{t}_0 = 5.7$)	Nei formulation	Rothman–Adams formulation ($c = 0.4$)
Ayoreo	0	0	0
Baniwa	1.72	2.00	2.60
Cayapo	1.37	1.50	1.64
Guaymí	0.70	0.08	0.83
Kraho	1.83	5.08	2.19
Macushi	1.48	0.92	1.76
Makiritare	1.46	1.61	1.75
Panoa	0.51	0.09	0.61
Piaroa	1.49	2.15	1.78
Wapishana	2.80	2.62	3.35
Xavante	3.69	4.14	4.40
Yanomama	0.12	0.04	0.14
Mean	1.43	1.69	1.71
		Av = 1.61	

The procedures were applied to the data of Tables 1 and 2, expressed as mutations/10^5 gametic loci/generation. The genetic polymorphisms of wide distribution have not been used in arriving at these estimates.

DISCUSSION

The estimates of Table 3 apply only to the rate at which mutation results in a variant detectable by the electrophoretic conditions used in this laboratory. For at least three reasons, it is clear that we are failing to identify the consequences of all mutational events. First, in most of our studies, only a single set of electrophoretic conditions was used for each protein, and experience has abundantly demonstrated that no one set of conditions will demonstrate all charge-change variants. Second, the presence of well-known genetic polymorphisms in some of the proteins studied interferes with the detection of genetic variants. Third, a detailed analysis of the triplet codons for a polypeptide such as hemoglobin has shown that random mutation would be expected to result in approximately twice as many electrophoretically silent amino acid substitutions (no charge change) as electrophoretically detectable variants (19–22). It would thus be conservative to adjust the present estimate by a factor of approximately 3, to 4.7×10^{-5}/locus per generation, to encompass all mutations resulting in amino acid substitution.

On the basis of a lesser body of data and more approximate methods than those used in the present paper, Neel (1) suggested that, if the polymorphisms of worldwide distribution were omitted from the calculation, the rate at which mutation resulted in protein variants (charged and uncharged) in Amerindians was 6×10^{-5}/locus per generation. The agreement with the present estimates is more than satisfactory.

Our present estimate does not yet include provision for synonymous mutations or mutations resulting in complete loss of enzyme activity. From the data of Mukai and Cockerham (23) on *Drosophila*, the latter (commonly termed "nulls") may be 5 to 6 times as frequent as those resulting in a detectable electrophoretic change, but these null phenotypes can result from mutation of either a structural or a controller/regulator gene. The small but detailed study by Adetugbo *et al.* (24) on spontaneous mutants of immunoglobulin in mouse myeloma cell lines clearly indicates the high potential for null mutants in structural loci in a mammal, and Siciliano *et al.* (25) have indicated how somatic cell hybridization may in the future be used with human material to determine what proportion of nulls results from mutation at a structural locus. For the present, however, we will not attempt to correct our estimate to include nulls, noting only that such an inclusion might substantially increase the estimate.

Several cautions must be raised in connection with these estimates. All the alleles contributing to the calculation are assumed to be neutral in their phenotypic effects. Appreciable amounts of selection could bias the estimates in either direction, depending on the nature of the selection. The error in the estimation of average tribal N could be considerable, an overestimate biasing the estimate of μ downward, and vice versa for an underestimate. Rothman and Adams (6) have mentioned that recent population growth can bias the estimate of μ downward. In this connection it is noteworthy that the two tribes that have almost surely expanded their numbers in recent generations, the Yanomama and the Guaymí, yielded two of the three lowest estimates of μ. Likewise, the error in the estimate of $\bar{t}_{0_A}$ could be considerable, although it seems unlikely to be off by a factor of 2. Finally, it would be premature to generalize from this estimate from Amerindians to all human populations; the possibility of regional/ethnic differences in mutation rates in man cannot be excluded on the basis of any present data.

Recently, Thompson and Neel (10) calculated the probability of a mutant gene achieving the numbers characteristic of a polymorphism in these Amerindian tribes, under various assumptions concerning the value of m (see above). Neel and Thompson (10) then estimated the mutation rate consistent with the observed number of electrophoretically demonstrable private polymorphisms listed in Table 2, for values and fluctuations of m consistent with the facts concerning Amerindian populations. Assigning values of m in this fashion is tantamount to an assumption of phenotypic neutrality on the part of the alleles responsible for these polymorphisms. The estimate is 0.7×10^{-5}/locus per generation. This estimate is subject to many of the same sources of error mentioned earlier and in addition is biased downward by the fact that, due to incomplete sampling, all the private polymorphisms within each tribe may not have been detected. Given these facts, the agreement with the earlier estimate based on all the electrophoretic variants, of 1.6×10^{-5}/locus per generation, is seen as satisfactory.

This work was supported in part by Department of Energy Contract EY-77-C-02-2828.

1. Neel, J. V. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3311–3315.
2. Stevenson, A. C. & Kerr, C. B. (1967) *Mutat. Res.* **4**, 339–352.
3. Cavalli-Sforza, L. L. & Bodmer, W. F. (1971) *The Genetics of Human Populations* (Freeman, San Francisco).
4. Vogel, F. (1975) in *Advances in Human Genetics*, eds. Harris, H. & Hirschhorn, H. S. (Plenum, New York), Vol. 5, pp. 223–318.
5. Nei, M. (1977) *Am. J. Hum. Genet.* **24**, 225–232.
6. Rothman, E. D. & Adams, J. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 5094–5098.
7. Neel, J. V. (1978) *Am. J. Hum. Genet.*, in press.
8. Neel, J. V. & Weiss, K. M. (1975) *Am. J. Phys. Anthropol.* **42**, 25–52.
9. Neel, J. V. & Thompson, E. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1904–1908.
10. Thompson, E. A. & Neel, J. V. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1442–1445.
11. Li, F. H. F., Neel, J. V. & Rothman, E. D. (1978) *Am. Nat.* **112**, 83–96.
12. Chagnon, N. A., Neel, J. V., Weitkamp, L., Gershowitz, H. & Ayres, M. (1970) *Am. J. Phys. Anthropol.* **32**, 339–349.
13. Migliazza, E. C. (1972) Dissertation (University of Indiana, Bloomington, IN).
14. Ward, R. H. & Neel, J. V. (1976) *Genetics* **82**, 103–121.
15. Neel, J. V. (1978) *Annu. Rev. Genet.*, in press.
16. Young, P. E. (1971) *Ngawbe. Tradition and Change Among the Western Guaymí of Panama* (University of Illinois Press, Urbana, IL).
17. Kimura, M. & Ohta, T. (1969) *Genetics* **63**, 701–709.
18. Thompson, E. A. (1976) *Am. J. Hum. Genet.* **28**, 442–452.
19. Shaw, C. R. (1965) *Science* **149**, 936–943.
20. Drake, J. W. (1970) *The Molecular Basis of Mutation* (Holden-Day, San Francisco).
21. Nei, M. & Chakraborty, R. (1973) *J. Mol. Evol.* **2**, 323–328.
22. Marshall, D. R. & Brown, A. D. H. (1975) *J. Mol. Evol.* **6**, 149–163.
23. Mukai, T. & Cockerham, C. C. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2514–2517.
24. Adetugbo, K., Milstein, C. & Secher, D. S. (1977) *Nature (London)* **265**, 299–304.
25. Siciliano, M. J., Bordelon, M. R. & Kohler, P. O. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 936–940.