Am J Hum Genet 38:649–666, 1986

A Revised Indirect Estimate of Mutation Rates in Amerindians

J. V. NEEL,[1] H. W. MOHRENWEISER,[1] E. D. ROTHMAN,[1,2] AND J. M. NAIDU[1,3]

SUMMARY

We have previously raised the possibility that the mutation rate resulting in rare electrophoretic variants is higher in tribal/tropical-dwelling/nonindustrialized societies than in civilized/temperate-dwelling/industrialized societies. Here, we report the results of examining 11 additional proteins for the occurrence of rare electrophoretic variants in 10 Amerindian tribes, for a total of 8,968 determinations and a total of 17,648 locus tests. When these data are combined with the results of all our previous similar studies of Amerindians, a total of 272,298 polypeptides, the products of 43 different loci, have been examined for the occurrence of rare electrophoretic variants. On the assumption that these variants are maintained by mutation pressure and are essentially neutral in their phenotypic effects, we have calculated by three different approaches that it requires an average mutation rate of 1.3×10^{-5}/locus per generation to maintain the observed variant frequency. Concurrently, we are reporting elsewhere that a direct estimate of the mutation rate resulting in electromorphs in various studies of civilized industrialized populations is 0.3×10^{-5}/locus per generation. Although this difference appears to have statistical significance, the nonquantifiable uncertainties to both approaches are such that our enthusiasm for a true difference in mutation rates between the two types of populations has diminished. However, even the lower of these estimates, when corrected for all the types of genetic variation

Received September 10, 1985.

This study was supported by U.S. Department of Energy Contract AC-02-82-ER-60089.

[1] Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109.

[2] Department of Statistics, University of Michigan College of Literature, Sciences and the Arts, Ann Arbor, MI 48109.

[3] National Scholar of the Government of India. Present address: Department of Anthropology, Andhra University, Waltair, India.

 0002-9297/86/3805-0004$02.00

that electrophoresis does not detect, implies total locus and gametic mutation rates well above those which in the past have dominated genetic thinking.

INTRODUCTION

For the past dozen years, we have been pursuing the possibility that mutation rates are substantially higher among unacculturated/tropical-dwelling/tribal populations than among civilized/temperate-dwelling/industrialized populations [1–3]. The issue is nontrivial, bearing as it does both on the question of the genetic and/or environmental control of human mutation rates and the impact on the current mutation rate (or absence thereof) of the various environmental changes that accompany civilization-industrialization. The basic data result from surveys for electrophoretic variants of serum proteins and erythrocyte enzymes. The evidence for higher mutation rates in tribal populations has all been indirect; the assumptions necessary to utilize this evidence are subject to error, the magnitude of which has been difficult to estimate. Recently, we have attempted a rather rigorous examination of the nature of these errors and their cumulative impact on the estimates [4]. We have also recently presented an updated direct estimate of the mutation rate to electromorphs in civilized populations (Caucasoids and Mongoloids), an estimate higher than the previous estimates [5]. Here, we will first present a substantial addition to the body of data available for indirect estimates of mutation rates in Amerindians. Then, using both the additional and the previous data, we will derive revised indirect estimates of the average mutation rates in the Amerindian populations studied and compute the errors to be attached to these estimates. This rate, although lower than the previous estimate, is still about four times higher than the direct estimate based on studies of the above-mentioned Caucasoid and Mongoloid populations; the difference is of borderline statistical significance. The new indirect estimate, thus, can be seen as reinforcing the results of the direct approach even while leaving open the possibility of somewhat higher mutation rates in tribal/tropical peoples. The important implications of these protein rates for nucleotide mutation rates will be explored.

DATA

Indirect estimates of mutation rates for electromorphs are based on the frequency of occurrence in population isolates of rare variants and "private" polymorphisms, such as might be maintained by mutation pressure. "Public" polymorphisms do not enter into the calculation. Our new data on such variants fall into two categories, one set resulting from the examination of up to 11 additional proteins not included in previous studies of the tribes in question, the other from an update on 26 previously investigated proteins. These 37 proteins are the products of 43 loci.

Data on an Additional 11 Enzymes in 10 Previously Studied Tribes

Table 1 lists the 11 additional proteins investigated and the tribes studied. Descriptions of the tribes will be found in previous publications from this laboratory. The

TABLE 1

THE RESULTS OF SURVEYING 10 AMERINDIAN TRIBES FOR THE OCCURRENCE OF RARE (NONPOLYMORPHIC) VARIANTS OF 11 ENZYMES

	ENZYME											
TRIBE	PEPC	PEPD	ESB	GLO	GOT	GPT	ALDHR	FUM	G6PD	ITPAS	PPAS	TOTAL
Baniwa	0/49	0/49	0/371	0/13	0/48	0/49	0/49	0/49	0/48	0/48	...	773
Guaymi	0/286	0/286	0/405	0/319	0/119	0/117	0/116	0/118	0/122	0/118	...	2,006
Kanamari	0/19	0/19	0/75	0/19	0/19	0/19	0/19	0/19	0/19	0/19	...	246
Macushi	0/252	0/251	0/68	0/216	0/68	0/62	0/64	0/66	0/68	0/65	...	1,180
Makiritare.....	0/55	0/56	0/79	0/77	0/61	0/57	0/30	0/58	0/51	0/47	...	571
Pano.....	0/35	0/35	0/319	0/27	0/39	0/35	0/35	0/35	0/38	0/35	...	633
Piaroa	0/15	0/15	0/15	0/16	0/17	0/15	0/11	0/15	0/17	0/15	...	151
Ticuna.....	0/1,762	0/1,762	0/1,263	0/1,762	0/108	0/103	0/105	0/104	0/104	0/102	0/1,742	8,917
Wapishana	0/48	0/48	0/49	0/47	0/49	0/46	0/38	0/48	0/33	0/47	...	453
Yanomama....	0/96	0/98	0/441	0/420	0/470	0/465	0/101	0/464	0/76	0/96	...	2,727
Total.....												17,657

proteins peptidase C (PEPC), peptidase D (PEPD), esterase B (ESB), glutamic oxaloacetic transaminase (GOT), and glucose-6-phosphate dehydrogenase (G6PD) were typed as described by Neel et al. [6]. Glyoxalase (GLO), glutamic pyruvate transaminase (GPT), and fumarase (FUM) were typed as described by Long et al. [7]. Adenosine triphosphatase (ITPAS) was typed as described by Vanderheiden [8] and Harris and Hopkinson [9], while acetaldehyde dehydrogenase (ALDH) was typed as described by Marada et al. [10]. For most of these determinations, an approximately 10% sample from the earlier collections from each tribe was typed. To the extent possible, the 10% sample was so chosen that only a single child per family was studied, and samples were available from both parents if needed. Table 1 presents the findings with respect to rare (i.e., nonpolymorphic) variants. Table 2 presents the findings with respect to the common genetic polymorphisms known from the reports of others to involve these proteins in most populations studied to date.

The well-known, worldwide polymorphisms of GLO and GPT (phenotypes 1, 1-2, and 2) were encountered in all tribes, the allele frequencies varying widely but no more widely than commonly observed for other polymorphic loci in Amerindian tribes. Two GOT variants were encountered; they were electrophoretically similar to the previously described type 2 (fast) and type 3 (slow) variants known to occur as polymorphisms in Mongoloid populations [11]. No differences could be detected between the GOT-2-type variants observed in the Yanomama and Pano tribes, or among the GOT-3-type variants occurring in five of the tribes, even when the five different buffer systems described in Wurzinger and Mohrenweiser [12] were employed in the electrophoretic procedures.

No rare variants occurring either in limited copies or as private polymorphisms were identified in 8,968 determinations. This is in contrast to our earlier experience with other systems, as summarized in 1978, in which the frequency of such variants was 2.8/1,000 [13]. Inasmuch as concurrent studies in other groups in this laboratory were yielding the usual frequency of rare variants, there is no possibility that the failure to detect any rare variants was due to technical factors. Given the manner in which indirect estimates of mutation rates are compiled (see below), the absence of rare variants in these systems will have an influence on the estimate of mutation rates that is quite disproportionate to the number of determinations involved.

An Update and Extension of the Data on the Previously Studied Systems

The results of our various studies on the types and frequency of rare variants in Amerindian tribes were last summarized in 1978 [13], and that summary was the basis for our last indirect estimate of human mutation rates [3]. Since that time, the results of five additional studies by ourselves on rare variants in Amerindians have been reported [6, 14–17]. The Kanamari and Ticuna covered in those reports had not been previously sampled, while the data on the Guaymi and Macushi represent new acquisitions for tribes studied previous to 1978. These five additional studies, plus the data being reported in the preceding section, increase the total number of determinations from this laboratory available for an estimate of the average mutation rate from 180,578 to 272,298.

Before this data base can be used for the intended calculations, two addenda to the previously reported findings are necessary. We have conducted extensive electrophoretic comparisons, using multiple buffer systems, of a number of apparently similar variants found in two or more tribes (GPI $2_{CAY\ 1}$, TF D_{CHI}, G6PD C, CP $A_{CAY\ 1}$), as well as of the GOT variants mentioned earlier. Only one variant has been subdivided. A CP variant observed in the Macushi and Wapishana tribes has been found to be electrophoretically distinct from the variant identified in the Cayapo and Xavante and designated CP $A_{CAY\ 1}$. The new variant is designated CP $A_{WAP\ 1}$ [18]. An additional classificatory change from the former data base involves the higher frequency of the $PGM_1\ 10_{MAC}$ variant among the new acquisitions from the Macushi tribe. This variant has now reached a frequency such that it must be classified as a "private" polymorphism.

TABLE 2

ALLELE FREQUENCIES FOR FOUR GENETIC POLYMORPHISMS ENCOUNTERED IN ONE OR MORE OF 10 AMERINDIAN TRIBES

	ENZYME								
	GLO			GPT			GOT		
TRIBE	No.	Allele	Frequency	No.	Allele	Frequency	No.	Allele	Frequency
Yanomama...	420	1	.238	465	1	.396	470	1	.974
		2	.742		2	.604		2	.026
Pano.....	37	1	.444	35	1	.571	39	1	.923
		2	.556		2	.429		2	.077
Ticuna.....	1,762	1	.832	103	1	.529	108	1	.977
		2	.168		2	.471		3	.023
Makiritare....	77	1	.701	57	1	.219	61	1	1.000
		2	.299		2	.718			
Piaroa	16	1	.312	15	1	.633	17	1	.941
		2	.688		2	.367		3	.059
Macushi	184	1	.288	62	1	.621	68	1	.963
		2	.712		2	.379		3	.037
Wapishana ...	47	1	.319	46	1	.554	49	1	.939
		2	.681		2	.446		3	.061
Guaymi.....	319	1	.668	117	1	.393	119	1	.992
		2	.332		2	.607		3	.008
Baniwa	13	1	.731	49	1	.561	48	1	1.000
		2	.269		2	.439			
Kanamari	19	1	.184	19	1	.395	19	1	1.000
			.816		2	.605			

The figure of 272,298 determinations mentioned above requires some explanation. We actually examined 37 different proteins for rare variants in one or more tribes, for a total of 198,773 determinations. However, LDH is the product of two loci, ESA as scored is the product of four loci, and hemoglobins A and A_2 are dimers with the added complication that the locus responsible for the α polypeptide of hemoglobins A and A_2 is duplicated. We, thus, were examining the products of 43 different genetic loci. Since, except for the sex-linked locus G6PD, the locus products reflected the presence of two alleles, the number of allele tests was 544,038. Two private polymorphisms, of serum albumin (ALB_{YAN2}) in the Yanomama and of acid phosphatase (ACP_{TIC}) in the Ticuna, contribute heavily to the frequency of rare variants, and, although restricted to single tribes, have polymorphic frequencies in the total Amerindian material. If these two variants are included in the calculation, our current (updated) frequency of so-called rare variants is 4.7 per 1,000 determinations. Without them, the figure becomes 1.6/1,000.

A REVISED, INDIRECT ESTIMATE OF THE MUTATION RATE FOR ELECTROPHORETIC VARIANTS IN AMERINDIANS

At this juncture, the use of the term "private" polymorphism must be examined more closely. It is customary in calculations of this type, following suggestions of Neel [1] and Nei [19], to exclude polymorphisms of widespread distribution from entering into the calculation. Tribal populations, however, frequently exhibit a type of genetic variation only rarely encountered in national, industrialized populations, stemming as the latter do from an amalgamation of numerous tribes hundreds and thousands of years ago. These are "private" polymorphisms, alleles that attain polymorphic frequencies within one or several closely related tribes but are not encountered in neighboring tribes. Such variants undoubtedly occurred in some of the tribes that were the precursors of the various national, industrialized populations of the world, but through the blending of tribal ancestries have been "diluted" to nonpolymorphic frequencies. We include such variants in the total count of rare variants in tribal populations (even though within a given tribe they are not rare), on the thesis that such variants are unavoidably commingled with the variants of civilized populations.

Since in fieldwork one presumably never samples all the individuals in the adult breeding population of a tribe (N), one may not detect some variants present in the tribal adults. We [20] have suggested an explicit adjustment of the data for this fact. The data necessary to this adjustment are given in tables 3 and 4. Let n be the average number of tests of adults for variants per enzyme per tribe and N be the number of adults in the tribe (i.e., the breeding population). Let k be the average number of variants (other than generally distributed polymorphisms) per enzyme determination per system (i.e., the sum of the entries in columns 3 and 4 of table 4 divided by the entry in column 4 of table 3). Then, the expectation of k, given the proportion of alleles, $g(i)$, that occur exactly i times in the population, is

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

TABLE 3

PERCENTAGE SAMPLED FROM EACH OF THE 14 TRIBES USED AS A BASIS FOR THE CALCULATION OF MUTATION RATES

Tribe	Estimated total population	Approximate *N* (adult generation)	No. locus products examined	Total determinations	Average no. sampled/ locus	Estimated no. adults sampled/ polypeptide (*n*)	% *N* sampled
Ayores.....	1,500	720	31	6,214	200	96	13.3
Baniwa	1,500	720	41	12,457	304	146	20.3
Cayapo	1,500	720	32	17,325	541	260	36.1
Guaymi.....	30,000	14,400	41	25,746	628	301	2.1
Kanamari	800	384	41	3,346	82	39	10.2
Kraho	600	288	32	6,109	191	92	31.9
Macushi	4,000	1,920	41	22,891	558	268	14.0
Makiritare.....	1,500	720	41	16,141	394	189	26.3
Panoa.....	18,000	8,640	41	11,018	269	129	1.5
Piaroa	3,000	1,440	38	4,237	112	54	3.8
Ticuna.....	15,000	7,200	42	59,114	1,407	675	9.4
Wapishana	2,000	960	42	20,131	479	230	24.0
Xavante.....	1,700	816	17	5,152	303	145	17.8
Yanomama....	15,000	7,200	42	62,417	1,486	713	9.9
				272,298			

TABLE 4

SUMMARY OF TYPES OF VARIANTS DETECTED BY ELECTROPHORESIS OF THE PRODUCTS OF LOCI, IN ADULTS OF 14 AMERINDIAN TRIBES, TOGETHER WITH THE ESTIMATES OF $\hat{K}$ AND THE CORRESPONDING MUTATION RATES CALCULATED BY THREE DIFFERENT APPROACHES

TRIBE	WIDELY DISTRIBUTED POLYMORPHISMS	PRIVATE POLYMORPHISMS	RARE VARIANTS	k	$\hat{K}$	AVERAGE MUTATION RATE $\times 10^{-5}$		
						Kimura-Ohta	Nei	Rothman-Adams
Ayores	3	0	0	0.000	0.000	0.00	0.00	0.00
Baniwa	8	1	1	0.049	0.105	1.28	1.59	1.53
Cayapo	5	1	1	0.063	0.099	1.20	1.33	1.44
Guaymi	10	1	1	0.049	0.598	0.36	0.05	0.43
Kanamari	7	0	1	0.024	0.082	1.87	*	2.23
Kraho	4	1	0	0.031	0.053	1.60	4.41	1.91
Macushi	8	2	2	0.098	0.264	1.21	0.76	1.44
Makiritare	7	0	2	0.049	0.091	1.11	1.28	1.32
Pano	8	1	0	0.024	0.409	0.41	0.07	0.50
Piaroa	9	1	0	0.026	0.192	1.17	5.86	1.40
Ticuna	9	1	0	0.024	0.084	0.10	0.03	0.12
Wapishana	9	2	5	0.167	0.327	2.99	2.85	3.57
Xavante	2	0	2	0.118	0.274	2.94	3.40	3.52
Yanomama	8	1	0	0.024	0.0812	0.10	0.03	0.12

* Negative value.

where K is the average number of variants per locus in that tribe. A geometric series model describing the number of copies of a mutant allele transmitted to the next generation has been fitted to the observed data [21] and, then, in conjunction with assumptions given in Rothman and Adams [20], used to obtain estimates of the sequence $g(i)$. Next, replacing the conditional expectation in the above formula with the observed number of alleles and substituting our estimates, $\tilde{g}(i)$, an extrapolation to $\hat{K}$ is obtained, namely,

$$\hat{K} = \frac{k}{1 - \Sigma\, \tilde{g}(i)\binom{2N-i}{2n}\Big/\binom{2N}{2n}} \,. \tag{2}$$

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

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

We will employ this method in our present paper. Columns 5 and 6 of table 3 present, by tribe, values for k and K. These values are averages, the number of polypeptides involved in the estimate per tribe ranging from 31 to 42 (table 3).

The indirect estimates (more properly, identities) for deriving mutation rates all require the following assumptions: (1) an equilibrium population, constant in its numbers for many generations, and with no in-migration from surrounding populations for the same period of time; (2) selective neutrality of the phenotypes associated with the alleles detected by electrophoresis; (3) an "infinite alleles" model, that is, mutation always resulting in a "new" allele with a distinguishable gene product; and (4) ultimate loss from the population of the alleles contributing to the calculation. Some or even all of these assumptions are undoubtedly incorrect for each of the tribes contributing to the dataset under consideration. Elsewhere we have examined in detail the probable nature of the departures from these assumptions and concluded that they are largely offsetting if one bases the estimate on a sufficiently large number of tribes that have been selected insofar as possible to meet the requirements of the model (4).

Mutation rates will be calculated in three different ways. Inasmuch as we have recently compared and contrasted these approaches in some depth [22], we will content ourselves here with a simple presentation of the identities.

1. *The Identity of Kimura and Ohta* [22]

$$\mu = \frac{K}{2N} \cdot \frac{1}{t_o} \,, \tag{4}$$

in which K and N are as previously defined and t_o is the average mutant survival time in generations for those mutations not going to fixation. Kimura and Ohta [22] did not discuss the estimation of K in the event N was incompletely sampled (which will generally be the case in fieldwork). Although Kumura and Ohta [22] presented a formula for calculating t_o, it depends upon a number of assumptions concerning population structure that are demonstrably unrealistic for tribal Amerindians. We have accordingly derived t_o on the basis of a Monte Carlo-type simulation of four Yanomama villages [21]. Since the N upon which the census of variants is based is the number in the adult generation, the value of t_o relevant to the calculation is for newly arisen mutations whose bearer survives to adulthood; this t_o is 5.6 generations.

2. *The Identity of Nei* [19], *Rewritten in the Format of Equation* (4)

$$\mu = \frac{k}{2N} \cdot \frac{1}{2 \log_e 2nq} , \tag{5}$$

where q is the prescribed threshold defining the notion of a rare allele. Nei recommends $q = .01$ and suggests that his formula(s) be used only when n is large (the second line of Nei [p. 227] contains a typographical error).

This rather attractive formula of Nei requires that to obtain an estimate of the mutation rate one provide only an estimate of the number of adults in a tribe and a count of the average number of rare alleles per locus in a sample. Neither the difficulty of extrapolating from k to K nor the estimation of t_o are required as in the Kimura-Ohta estimator. Furthermore, since the right side of equation (5) must remain constant whatever the value of n, provided n is large, we can relate equation (5) to equation (4) by setting n to N and k to K. In this form,

$$\mu = \frac{K}{2N[2 \log_e 2Nq]} . \tag{6}$$

Relating this equation to Kimura and Ohta's formulation shows that

$$t_o = 2 \log_e 2Nq . \tag{7}$$

Now with $q = .01$, a value recommended by Nei, we see a marked difference between the Kimura-Ohta value of t_o and the implicit value of this parameter in the Nei formula. The difference, expressed numerically, is given by t_o(Nei) $= t_o$(Kimura-Ohta) $- 9.2103$. In particular with $N = 7{,}500$, we find t_o(Nei) $= 5.4161$ as compared with t_o(Kimura-Ohta) $= 14.6264$, more than a twofold difference. In the usual situation in which a subsample of n adults is drawn from the total N, Nei's formula uses k in the numerator while the Kimura-Ohta formula requires an extrapolated estimate of K. So for certain special cases of n and N, the substantial difference in the denominator may be counterbalanced yielding roughly equal estimates of the mutation rate. To see when both esti-

mates will yield comparable values, we must have the estimated K of Kimura-Ohta

$$K = \frac{k \log_e 2Nq}{\log_e 2nq} \quad . \tag{8}$$

3. *The Identity of Rothman-Adams* [20]

The Rothman-Adams estimator is based on a balance, at equilibrium, between the expected number of alleles entering, through mutation and drift, the class of alleles with exactly one copy and those alleles leaving this class in a single generation. This estimator may be expressed in the Kimura-Ohta form

$$\hat{\mu} = \frac{\hat{K}}{2N} [\tilde{g}(1) = \Sigma \tilde{g}(i) P_{i1}] \quad , \tag{9}$$

where the P_{i1}, $i = 1, 2, \ldots 2N$ denote the probability of an allele with i copies transmitting only one copy to the next generation. Estimates of these probabilities, as for the $g(i)$ sequence, are obtained from a geometric series model fitted to data of Li et al. [21].

The quantity in brackets in equation (9) is, at equilibrium, a measure of the rate of loss of alleles per generation. Equating equation (9) with equation (4) gives

$$t_o = [\tilde{g}(1) - \Sigma \tilde{g}(i) P_{i1}]^{-1} \quad . \tag{10}$$

Lack of agreement between the estimates of the right and left side of equation (10) could suggest a departure from equilibrium.

Table 3 presents the basic data concerning the estimated proportion of the adult breeding generation (N) that was sampled. The term "adult" encompasses all persons aged 15–40. Variants not encountered in the adult generation as defined do not enter into the calculation. The number of rare variants and private polymorphisms observed to be unique to each tribe and the indirect estimates of mutation rate to which they lead are presented in table 4. As noted, the value of t_o substituted in the Kimura-Ohta identity is 5.6. It is specific to a population with Amerindian breeding structure. The unweighted average of the tribal mutation rate estimates for electromorphs yielded by the Kimura-Ohta approach is 1.2×10^{-5}/locus per generation and by the Rothman-Adams approach, 1.4×10^{-5}/locus per generation. A problem arises with the Nei estimator when n is small. Indeed, when $2n < 1/q$, the estimate is a negative number (i.e., with $q = .01$, a negative estimate results if $n \leq 50$). Nei advises that his estimator be used only with n large. When we omit the one case in our series with $n < 50$, the average of the remaining 13 estimates is 1.7×10^{-5}/locus per generation. Unfortunately, a number of other sample sizes are also probably inappropriately small for the use of the Nei estimator. We will conclude only that the results of the Nei approach, in general, confirm the results

of the other two but should not be used further. The average of the results of the other two estimates based on all the data is 1.3×10^{-5}/locus per generation. This is to be contrasted with our last estimate, of 1.6×10^{-5}/locus per generation [2]. The decrease is largely the result of the failure to detect any additional rare variants in the 11 systems for which data are reported in this paper, plus the paucity of variants in the Ticuna [6].

Elsewhere we have argued that each mutation resulting in an electrophoretic variant of a protein corresponds roughly to two additional undetected mutations involving nucleotide substitutions [23]. If we assume that the average polypeptide on which this estimate is based requires 1,000 nucleotides for its specification, and that exon rates are characteristic of the DNA as a whole, this finding implies a nucleotide mutation rate/generation in this population of 3.9×10^{-8}. While this on the one hand represents an amazing fidelity of replication, with at least 2×10^{9} nucleotides in the human haploid genome, the average gamete of the populations under study should contain 78 mutations of the nucleotide substitution type. This estimate would, of course, not include mutations resulting in small deletions, duplications, or rearrangements of the genetic material. Even if this estimate were too high by a factor of 2, such a total rate would raise a number of serious questions for the theory of population genetics, questions which we propose to treat elsewhere.

ASSIGNMENT OF AN ERROR TO THE ESTIMATES

The standard error of the mean, computed as $S/\sqrt{T}$, where S is the standard deviation among the tribal estimates, and T, the number of tribes, provides a first approximation to the standard deviation of the estimates. The resulting value based on the estimates yielded by the Kimura-Ohta approach is 0.3×10^{-5}, and the same value is obtained for the Rothman-Adams estimates. Pollak [24] recently derived the error term to be attached to the Nei estimator. Although we have above indicated concern that our sample sizes are too small for the appropriate application of the Nei estimator, it is relevant that the Pollak derivation indicates that the error to be attached to the 13 positive estimates derived by the Nei approach is 0.4×10^{-5}.

Elsewhere we discussed the possibility of systematic departures from the assumptions necessary for the calculations and have concluded that the most probable result of such departures is an underestimate of the true mutation rate [4]. In particular, we doubt that all of the variants that enter into this calculation can be neutral in their effect on fitness. Such systematic departures should not influence the magnitude of the error term. On the other hand, individual tribes may depart in a nonsystematic way from the assumptions necessary to the calculation. The assumptions most likely to be violated are of constant population size and absence of genetic exchange with neighboring tribes. A recent increase in population size should result in an underestimate of the mutation rate, whereas genetic exchange with neighboring tribes should result in an overestimate. It is, in this connection, noteworthy that two of the three smallest estimates are yielded by tribes that have almost certainly expanded their numbers significantly this century (the Ticuna and Yanomama), whereas one of

the two highest estimates is based upon a tribe (the Wapishana) that is alleged to have absorbed the remnants of several other tribes during the present century. Although these departures are compensating in their effect upon the mean, they should inflate the error term. Unfortunately, no correction is possible, and we will use the present error term as the best available.

DISCUSSION

Other Indirect Estimates of Human Mutation Rates

To date, there have been three other efforts to apply the indirect approach to human mutation rates to data on electromorphs in tribal populations. Unfortunately, in each instance, the populations under consideration depart from the assumptions of the model even further than Amerindian tribes.

(1) Bhatia et al. [25], using the excellent data of Kirk et al. on Australian Aborigines, calculated a rate for this latter group by the three methods used in this paper. Unfortunately, tribal population numbers have been dramatically reduced in the past several hundred years, and tribal boundaries undoubtedly blurred, with increased tribal intermarriage. This possibility is borne out by the single set of tribal data they present, on the Waljbiri, who share four of their five rare variants with the remaining Aboriginal population; we do not believe this tribe is a suitable base for an estimate. They also present an estimate derived from the total present Aboriginal population. Basing their estimate on the current population size should bias the estimate upwards, since the mutations arose when the population base was larger. On the other hand, using the total population of Aborigines as the basis for their estimate will bias it downwards, since the effective breeding size of the population within which the observed rare variants and private polymorphisms arose and established themselves was presumably much smaller than the total population. Some idea of the magnitude of this latter bias can be gained from an empiric demonstration by Neel et al. [26]; the average tribally based estimate of 1.6×10^{-5}/locus per generation for electromorphs was reduced to 0.7×10^{-5} when the total estimated adult tribal breeding population of all South Amerindians was used as the population base for the estimate. Be this as it may, when the estimate of variant number was corrected by the method of Neel and Rothman for the fact that only a fraction of the Aboriginals had been sampled, the data yielded the following three estimates: Kimura-Ohta, 1.0×10^{-5}; Nei, 0.6×10^{-5}, and Rothman-Adams, 1.3×10^{-5}. Since we believe the greater bias is associated with treating the Aboriginals as one interbreeding population, we find this result in satisfactory agreement with the current result.

(2) Bhatia et al. [27], again using the data of Kirk et al., also calculated by the indirect method mutation rates for the populations of Papua New Guinea. The rates estimated on the assumption that linguistic groups constitute the breeding unit, with I corrected as suggested by Rothman and Adams [20], were: Kimura-Ohta, 0.2×10^{-5}; Nei, 0.2×10^{-5}; Rothman-Adams, 0.6×10^{-5}. The first two rates are calculated on the unrealistic assumption that mutant loss occurs as in the idealized population structure assumed by Kimura and Ohta [22] (see [21]

for discussion of how tribal populations depart from this structure). Correction for this would render the first two estimates much more similar to the third. We suggest that serious consideration must be given to the possibility that a rapid expansion of the Papua New Guinea population during the last 300 years is largely responsible for the difference between the estimate of 0.6×10^{-5} given above and the estimate of this paper of 1.3×10^{-5} for Amerindians.

(3) Chakraborty and Roychoudhury [28] also presented an indirect estimate of mutation rate for the Kadars of Kerala State, India. We have discussed elsewhere why this is an inappropriate population [3].

Concordance of the Current Estimate with the Results of the Thompson-Neel Approach

Above we referred to the occurrence of "private" polymorphisms in Amerindian tribes. Thus far, with the addition of the $PGM_1 10_{MAC}$ variant mentioned above, 12 examples of the phenomenon have been recognized [29]. These are presumed due to a founder effect resulting from the occurrence and survival of a single, unique mutation at some time in tribal history. Based on the results of a simulated 400-year span of the demography of four Yanomama villages [21], we computed the probability of founder effects of various magnitudes, using a branching process model [30]. From a consideration of the number of these polymorphisms in relation to the number of loci surveyed, and the number of alleles with which each polymorphism is represented, the approximate mutation rate necessary to maintain this array can be inferred. On the assumption of selective neutrality of the alleles in question, the mutation rate for electromorphs most consistent with the data was 0.7×10^{-5}/locus per generation [31]. The error associated with this estimate is undoubtedly relatively large but indeterminate.

A Comparison with the Results of Direct Estimates

We recently reported that in a series of 529,971 tests for spontaneous mutations detected electrophoretically in Japanese children and young adults, the rate was 0.6×10^{-5}/locus per generation, with a 95% confidence interval between 0.1×10^{-5} and 1.7×10^{-5} [5]. Inclusion in this estimate of all the data available in the world's literature on civilized populations resulted in a rate of 0.3×10^{-5}, with 95% confidence limits of 0.1×10^{-5} and 0.8×10^{-5} [5]. A comparison between the mutation rates in civilized and tribal populations would traditionally be based on the difference between the two estimates divided by the standard error of this difference. Unfortunately, in this instance, this statistic is not normally distributed under the null hypothesis of no difference between the mutation rates. Estimates of the skew and kurtosis are 0.02 and 3.80, respectively, while a standard normal density has no skew and a kurtosis of 3. Thus, the distribution of the statistic is slightly skewed and has heavier tails than a normal density. Using an Edgeworth expansion (cf. [32]), with parameters estimated from the data, we find the observed value of the statistic corresponds to a P-value $< .04$, while for the normal approximation a P-value of $< .01$ is obtained. Of course, the Edgeworth expansion for the

distribution of the statistic provides a suitable approximation only when the moments are known exactly, whereas here we estimated moments. On the other hand, the large sample sizes involved certainly imply that the difference between the mutation rates in these two populations are at least at the edge of statistical significance.

As mentioned in a preceding section, the magnitude of the "total" nucleotide rate suggested by the indirect approach raises a number of profound questions for classical population genetics. Even the lower rate suggested by the direct approach raises many of the same questions. We shall, for now, adopt the position that, at the very least, the results of the indirect approach applied to Amerindians confirm the magnitude of the nucleotide and total mutation rate suggested by the direct approach in nontribal populations, while at the same time, especially in view of the data to be discussed in the next section, still leaving open the possibility of higher rates in unacculturated tribal populations.

The Higher Frequency of Rare Variants in Tribal/Tropical Dwelling Populations

A comparison of the frequency of rare variants of 20 proteins examined in three different tribal and three different civilized populations revealed a higher frequency of these variants in tribal populations [3]. Despite the fact that no rare variants were encountered in the additional locus tests reported in this paper, that finding remains correct. As noted above, the rare variant plus private polymorphism frequency in this series now stands at 4.7/1,000 polypeptides examined. In a Caucasian series studied in London with respect to many of the same proteins, using essentially the same techniques, the frequency was 1.6/1,000 [33]. For Caucasoids in Ann Arbor, the corresponding figure is 2.4/1,000 (Mohrenweiser and Neel, manuscript in preparation), whereas for Japanese, the frequency was 2.0/1,000 [34, 35]. In another recent comparison, Naidu et al. [36] record that for a set of 12 enzymes, a subset of those surveyed in this study, the frequency of rare electrophoretic variants per 1,000 determinations was 2.8 for the long-civilized populations thus far studied in South India but 3.6 for the tribal populations of the same area. The difference, while not significant, is in the same direction noted in the data we have just discussed.

In two populations with identical breeding structures, the population with the higher mutation rate should support a greater number of different rare variants and a higher total frequency of these variants. The highly subdivided nature of tribal populations and the higher infant and childhood mortalities should result in a higher rate of mutant loss from these populations than from civilized populations. The fact is that, this notwithstanding, variant frequencies for electromorphs are higher in the tribal populations. One obvious possibility is that the estimated threefold mutation rate differences between the two populations, while not statistically significant, are valid.

It is pertinent to efforts to understand the significance of this difference that Mohrenweiser and Neel [37] found no apparent difference in the frequency of carriers of enzyme deficiency alleles in Amerindians from the findings in Caucasians [38] or Japanese [39]. This finding is consistent either with the

hypothesis of higher mutation rates in the tribal populations, with more efficient elimination of alleles responsible for grossly decreased enzyme activity, or a different distribution, in the two types of populations, of the selective coefficients associated with these two classes of variants, the mutation rates being the same in both populations.

Other Explanations of the Higher Frequency of Rare Variants in Tribal/Tropical Populations

If one elects to conclude that there is no difference between the mutation rates of the two types of populations under consideration, one then has to address the question of why there are greater frequencies of rare variants in these tribal populations. One possibility involves the requirement that the variants found among unacculturated populations are subject to greater variation in selective values than one may anticipate among the variants found in civilized populations. Such a requirement has been studied in the context of a model of weak selection [40] and the qualitative implication would be a higher frequency of rare variants. However, it is not clear whether the quantitative impact could account for the differences found.

REFERENCES

1. NEEL JV: "Private" genetic variants and the frequency of mutation among South American Indians. *Proc Natl Acad Sci USA* 70:3311–3315, 1973
2. NEEL JV, ROTHMAN ED: Indirect estimates of mutation rates in tribal Amerindians. *Proc Natl Acad Sci USA* 75:5585–5588, 1978
3. NEEL JV, ROTHMAN ED: Is there a difference between human populations in the rate with which mutation produces electrophoretic variants? *Proc Natl Acad Sci USA* 78:3108–3112 1981
4. ROTHMAN ED, NEEL JV: Indirect estimates of human mutation rates. Submitted for publication
5. NEEL JV, SATOH C, GORIKI K, ET AL.: The rate with which spontaneous mutation alters the electrophoretic mobility of polypeptides; implications for DNA rates. *Proc Natl Acad Sci USA* 83:389–393, 1986
6. NEEL JV, GERSHOWITZ H, MOHRENWEISER HW, ET AL.: Genetic studies on the Ticuna, an enigmatic tribe of Central Amazonas. *Ann Hum Genet* 44:37–54, 1980
7. LONG JC, NAIDU JM, MOHRENWEISER HW, ET AL.: Genetic characterization of Gainj and Kalam speaking people of Papua New Guinea. *Am J Phys Anthropol.* In press, 1986
8. VANDERHEIDEN BC: Genetic studies of human erythrocyte inosine triphosphatase. *Biochem Genet* 3:289–297, 1969
9. HARRIS H, HOPKINSON DA: *Electrophoretic Methods in Human Genetics*. New York, Elsevier, 1976
10. MARADA S, AGARWAL PD, GOEDDE WH: Isozyme variations of acetaldehyde dehydrogenase in human tissues. *Hum Genet* 44:181–185, 1978
11. CHEN S-H, GIBLETT ER: Genetic variation of soluble glutamic-oxaloacetic transaminase in man. *Am J Hum Genet* 23:419–424, 1971
12. WURZINGER KH, MOHRENWEISER HW: Studies on the genetic and non-genetic (physiological) variation of human erythrocyte glutamic oxaloacetic transaminase. *Ann Hum Genet* 46:191–201, 1982
13. NEEL JV: Rare variants, private polymorphisms, and locus heterozygosity in Amerindian populations. *Am J Hum Genet* 30:465–490, 1978

14. MOHRENWEISER HW, NEEL JV, MESTRINER MA, ET AL.: Electrophoretic variants in three Amerindian tribes: the Baniwa, Kanamari, and Central Pano of Western Brazil. *Am J Phys Anthropol* 50:237–246, 1979
15. MESTRINER MA, SIMOES AL, SALZANO FM: New studies on the esterase D polymorphism in South American Indians. *Am J Phys Anthropol* 52:95–101, 1980
16. BARRANTES R, SMOUSE PE, NEEL JV, ET AL.: Migration and genetic infrastructure of the Central American Guaymi and their affinities with other tribal groups. *Am J Phys Anthropol* 58:201–214, 1982
17. SALZANO FM, MOHRENWEISER HW, GERSHOWITZ H, ET AL.: New studies on the Macushi Indians of Northern Brazil. *Ann Hum Biol* 11:337–350, 1984
18. MOHRENWEISER HW, DECKER RS: Identification of several electrophoretic variants of human ceruloplasmin including $Cp^{Michigan}$, a new polymorphism. *Hum Hered* 32:369–373, 1982
19. NEI M: Estimation of mutation rate from rare protein variants. *Am J Hum Genet* 29:225–232, 1977
20. ROTHMAN ED, ADAMS J: Estimation of expected number of rare alleles of a locus. *Proc Natl Acad Sci USA* 75:5094–5098, 1978
21. LI FHF, ROTHMAN ED, NEEL JV: A second study of the survival of a neutral mutation in a simulated Amerindian population. *Am Nat* 112:83–96, 1978
22. KIMURA M, OHTA T: The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics* 63:701–709, 1969
23. NEEL JV: A revised estimate of the amount of genetic variation in human proteins: implications for the distribution of DNA polymorphisms. *Am J Hum Genet* 36:1135–1148, 1984
24. POLLAK E: On three methods for estimating rates indirectly. *Am J Hum Genet* 38:209–227, 1986
25. BHATIA KK, BLAKE NM, KIRK RL: The frequency of private electrophoretic variants in Australian Aborigines and indirect estimates of mutation rate. *Am J Hum Genet* 31:731–740, 1979
26. NEEL JV, MOHRENWEISER HW, HANASH S, ET AL.: Biochemical approaches to monitoring human populations for germinal mutation rates: I. Electrophoresis, in *Utilization of Mammalian Specific Locus Studies in Hazard Evaluation and Estimation of Genetic Risk,* edited by DESERRES CF, SHERIDAN R, New York, Plenum Press, 1983, pp 71–93
27. BHATIA KK, BLAKE NM, SERJEANTSON SW, KIRK RL: Frequency of private electrophoretic variants and indirect estimates of mutation rate in Papua New Guinea. *Am J Hum Genet* 33:112–122, 1981
28. CHAKRABORTY R, ROYCHOUDHURY AK: Mutation rates from rare variants of proteins in Indian tribes. *Hum Genet* 43:179–183, 1978
29. NEEL JV: Isolates and private polymorphisms, in *Population Structure and Genetic Disorders,* edited by ERIKSSON A, New York, Academic Press, 1980, pp 173–193
30. THOMPSON EA, NEEL JV: Probability of a founder effect in a tribal population. *Proc Natl Acad Sci USA* 75:1442–1445, 1978
31. NEEL JV, THOMPSON EA: Founder effect and number of private polymorphisms observed in Amerindian tribes. *Proc Natl Acad Sci USA* 75:1904–1908, 1978
32. KENDALL M, STEWART A: *The Advanced Theory of Statistics,* vol 1, 2nd ed. London, Charles Griffin, 1963
33. HARRIS H, HOPKINSON DA, ROBSON EB: The incidence of rare alleles determining electrophoretic variants: data on 43 enzyme loci in man. *Ann Hum Genet* 37:237–253, 1974
34. NEEL JV, UEDA N, SATOH C, ET AL.: The frequency in Japanese of genetic variants of 22 proteins. V. Summary and comparison with data on Caucasians from the British Isles. *Ann Hum Genet* 41:429–441, 1978
35. NEEL JV, SATOH C, HAMILTON HB, ET AL.: Search for mutations affecting protein

structure in children of atomic bomb survivors: preliminary report. *Proc Natl Acad Sci USA* 77:4221–4225, 1980
36. NAIDU JM, MOHRENWEISER HW, NEEL JV: A sero-biochemical study of Jalari and Brahmin caste populations of Andhra Pradesh. *Hum Hered* 35:148–156, 1985
37. MOHRENWEISER HW, NEEL JV: A "disproportion" between the frequency of rare electromorphs and enzyme deficiency variants in Amerindians. *Am J Hum Genet* 36:655–662, 1984
38. MOHRENWEISER HW: Frequency of enzyme deficiency variants in erythrocytes of newborn infants. *Proc Natl Acad Sci USA* 78:5046–5050, 1981
39. SATOH C, NEEL JV, YAMASHITA A, GORIKI K, FUJITA M, HAMILTON HB: The frequency among Japanese of heterozygotes for deficiency variants of 11 enzymes. *Am J Hum Genet* 35:656–674, 1983
40. ROTHMAN ED, WEBER NC: A model of weak selection in the infinite alleles framework. *J Math Biol.* In press, 1986

THE NATIONAL INSTITUTE OF GENERAL MEDICAL SCIENCES HUMAN GENETIC MUTANT CELL REPOSITORY. The National Institute of General Medical Sciences sponsors a bank of human cell lines representing a wide variety of genetic disorders. The bank, known as the Human Genetic Mutant Cell Repository, is maintained at the Coriell Institute for Medical Research in Camden, New Jersey. The purpose of the Repository is to promote and facilitate research on human genetic disease by providing to qualified investigators cell cultures that are well characterized, thoroughly documented, and free of contamination.

The Repository contains mainly fibroblast and lymphoblast cultures from individuals with a wide range of inherited metabolic disorders and chromosomal aberrations. Approximately 4,000 individual cultures, representing approximately 300 genetic diseases and 500 chromosomal aberrations, were listed in the 1985 catalog. In recent years, a special effort has been initiated to establish a collection of lymphoblastoid cultures from extended reference pedigrees for developing the human linkage map with respect to known DNA markers, as well as for mapping gene loci. To accomplish this objective, the Repository obtained a series of lymphoblastoid cultures from: six apparently normal multigeneration Utah families; an Old Order Amish pedigree that contains a high incidence of primary affective disorder; a Venezuelan pedigree that shows a high incidence of Huntington disease as well as a Venezuelan reference pedigree; 11 cystic fibrosis families, many with multiple affected children; 65 members of a family with maturity onset diabetes of the young; a tuberous sclerosis pedigree; a family with Gardner syndrome type III; and several large fragile X-linked mental retardation families.

A moderate fee is charged for the cell cultures, and a discount is available for large orders. Information about the cultures in the Repository and details of procedures for submitting and obtaining cell cultures may be obtained by contacting: Dr. Arthur E. Greene, The Human Genetic Mutant Cell Repository, Coriell Institute for Medical Research, Copewood and Davis Streets, Camden, NJ 08103. Telephone: (609)966-7377. Cable address: INMEDRES.