Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 9611–9615, December 1986
Genetics

Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism

(high-resolution maps/neutral mutation theory/mtDNA/genetic diversity/population structure)

THOMAS S. WHITTAM*, ANDREW G. CLARK*, MARK STONEKING†, REBECCA L. CANN†‡, AND ALLAN C. WILSON†

*Department of Biology, Pennsylvania State University, University Park, PA 16802; and †Department of Biochemistry, University of California, Berkeley, CA 94720

Communicated by James F. Crow, August 28, 1986

ABSTRACT **Restriction maps of 145 human mtDNAs representing samples from five geographic regions were used to construct multilocus genotypes for 28 genetic loci of the mitochondrial genome. Alleles were defined as distinct combinations of the presence or absence of polymorphic restriction sites within each locus. The 28 loci included 13 genes encoding proteins, 10 genes specifying tRNAs, 2 genes specifying rRNAs, and 3 noncoding regions consisting of the D loop, the light strand origin of replication, and the 5′ noncoding sequence. In 35 comparisons of allele frequency distributions to expected distributions predicted by neutral mutation theory (assuming an infinite alleles model), the results revealed that most genetic diversity values (71%) fell within the range predicted by the neutral model; however, excesses in the frequencies of common alleles and in the number of singleton alleles within populations were observed at specific loci. Departures from the neutral mutation model are most readily explained by the effects of the recent expansion of the human population and the action of purifying selection. Coefficients of population differentiation suggest that gene flow of mtDNA types between certain geographic regions may be limited.**

Data on the variability in mitochondrial DNA sequences have accumulated rapidly through the application of restriction endonuclease analysis to a variety of organisms (refs. 1–14 and references therein). Besides being used to construct phylogenetic trees, test hypotheses about the mechanisms of mtDNA evolution, and explore population structure within species (1–14), these data have encouraged the development of new statistical methods for estimating polymorphism and genetic diversity at the nucleotide level (15–18) and have inspired the derivation of population genetic theory that focuses on the evolution of organelle genes (19–26).

Despite these advances, many of the questions concerning the roles of selection, mutation, migration, and genetic drift in the evolution of organelle genes remain the same as those addressed for nuclear genes. In this report data on restriction site polymorphism in human mtDNA are translated into familiar population genetic terms—i.e., multilocus genotypes and allele frequency distributions. The approach is applicable to high-resolution maps such as those available for 145 individuals within the human species (12, 27). The construction of such maps was possible because the entire mtDNA sequence for one human being is known (28). The ability to map restriction sites into specific functional regions of the mtDNA genome—i.e., into loci—allows us to define alleles as distinct combinations of restriction sites observed within a locus.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

By interpreting patterns of site polymorphism as allelic variation at specific loci (29), one gains access to the statistical tests and traditional measures of population structure that have been developed to study genetic variation in natural populations. Additional advantages of this approach include the ability to detect sequence variation in nonstructural genes and the high resolution that yields a rich diversity of alleles necessary for robust statistical analysis. After we identify alleles (or nucleomorphs, see refs. 17 and 29) of mitochondrial loci in samples from human populations, we compare the allele frequency distributions to the predictions of a model of selectively neutral mutations. Although the mtDNA loci are completely linked and therefore share the same phylogenetic history, we treat each locus individually, because neutral mutations arise independently at separate loci, and functional constraints on gene products presumably vary independently across loci. Under conditions of complete linkage, however, the evolutionary fate of neutral mutations also depends on the occurrence of adaptive mutations anywhere in the mitochondrial genome. From the results of single-locus comparisons to the neutral mutation model we draw inferences about the past action of genetic drift, migration, and natural selection in determining the evolution of human mitochondrial genes.

MATERIALS AND METHODS

Data on restriction site polymorphism of 145 human mtDNA genomes were compiled from studies by Cann *et al.* (12, 30) and Stoneking *et al.* (27). Details of the high-resolution restriction mapping procedures are described elsewhere (12, 27). Briefly, the data consist of the presence or absence of 190 restriction sites produced by 11 restriction endonucleases in individuals originating from five broad geographic regions (I–V). Human samples include (I) 20 people of African descent; (II) 34 of Asian descent; (III) 21 Australian aborigines; (IV) 44 from Europe, North Africa, and the Middle East; and (V) 26 from Papua New Guinea. Length mutations, described in this same data set (5), have been omitted from the analysis, because they have not been mapped to specific loci.

The mtDNA genotype of each individual was specified by a unique 190-digit binary number, where each digit represents the presence or absence of a polymorphic restriction site. All sites were mapped on the human mtDNA sequence (12, 27) and were assigned to specific loci. We considered 28 loci as follows: (*i*) 13 structural genes that encode 2 ATPase subunits (A6 and A8); cytochrome oxidase subunits I, II, and III; cytochrome *b*; 7 NADH dehydrogenase subunits (1–6, 4L); (*ii*) 10 genes that specify tRNAs; (*iii*) 2 genes that specify rRNA subunits (12S and 16S); and (*iv*) 3 noncoding se-

‡Present address: Department of Genetics, University of Hawaii, Honolulu, HI 96822.

quences consisting of the D loop, the noncoding sequence adjacent to the light strand origin of replication, and the 5′ noncoding sequence (see ref. 14 for a map of the human mtDNA genome). By comparing the combinations of polymorphic sites across individuals, we identified alleles as distinct haplotypes or combinations of restriction sites falling within each locus. In this way, information on the presence or absence of 190 sites for an individual was interpreted as a multilocus genotype defined from the profile of alleles across 28 loci. Allele frequencies were then tabulated within each geographic region and in the pooled sample.

RESULTS

Levels of genetic variation detected within each of 28 loci of human mtDNA are summarized in Table 1. The number of alleles per locus is related to the size of the locus and to the number of polymorphic sites (see below). The average number of alleles per locus for the pooled sample of 145 mtDNA genomes is 8.8, and the range is from 40 for the D loop to 2 for several tRNAs. Genetic diversity for each locus was calculated for each geographic region as $h = (1 - \Sigma x_i^2)N/(N - 1)$ where N is the number of genomes sampled and x_i is the frequency of the ith allele (32, 33). The average genetic diversity within geographic regions, H_S, ranges from 0.850 for the D loop to 0.014 for four tRNAs, and the average over all loci is 0.215.

Table 1. Genetic variation for 28 loci of human mtDNA based on restriction maps of 145 individual genomes sampled from five geographic regions

Locus	Size, bp	Polymorphic sites, no.	Alleles, no.	Diversity within regions (H_S)	Differentiation between regions (G_{ST})
D loop	737	22	40	0.850	0.028
ND 5	1812	22	30	0.692	0.098
CytOxase I	1539	23	24	0.569	0.098
Cyt *b*	1140	12	14	0.480	0.147
ND 4	1377	11	14	0.291	0.040
ND 1	954	11	12	0.361	0.112
CytOxase III	783	10	11	0.183	0.005
CytOxase II	681	9	11	0.192	0.017
ATPase 6	678	9	11	0.312	0.025
16S rRNA	1559	8	11	0.252	0.093
12S rRNA	953	8	10	0.384	0.074
ND 2	1041	8	9	0.217	0.061
ND 6	525	5	6	0.068	0.007
ND 4L	294	4	5	0.082	0.000
ND 3	345	4	4	0.484	0.015
$tRNA^{Val}$	69	3	4	0.104	0.027
$tRNA^{Thr}$	66	3	4	0.081	0.000
LS origin	31	2	3	0.041	0.011
ATPase 8	204	2	3	0.041	0.000
5′ Noncoding	384	2	3	0.027	0.008
$tRNA^{Met}$	68	2	3	0.068	0.000
$tRNA^{Trp}$	68	2	3	0.028	0.000
$tRNA^{Pro}$	69	2	2	0.014	0.000
$tRNA^{Ser}$	72	2	2	0.014	0.000
$tRNA^{Leu}$	75	1	2	0.095	0.095
$tRNA^{Arg}$	65	1	2	0.067	0.000
$tRNA^{Glu}$	68	1	2	0.014	0.000
$tRNA^{Lys}$	70	1	2	0.014	0.000
Average	562	6.8	8.8	0.215	0.063

Alleles were defined by distinct combinations of restriction sites. Genetic diversity within regions is measured by H_S, the weighted average of diversities calculated within each region. The relative genetic differentiation among regions is measured by G_{ST}, the ratio of the among-region diversity ($H_T - H_S$) to the genetic diversity in the total sample (H_T). Average G_{ST} is calculated from the average values of H_S and H_T as described by Nei (31). ND, NADH dehydrogenase; bp, base pairs; CytOxase, cytochrome oxidase; Cyt, cytochrome; LS, light strand.

Testing Neutrality. The estimates of genetic diversity for each locus were compared to those predicted from the neutral theory of molecular evolution (34). Under the hypothesis of strict neutrality, the sampling theory of Ewens (35, 36) and Watterson (37) predicts the relationship between the number of alleles in a sample and single-locus identity, F (= $1 - h$). The theory assumes that each new mutation is to a novel allele (the "infinite alleles" model, ref. 38) and that the distribution of allele frequencies has achieved a steady state with respect to mutation and random genetic drift. Fig. 1 shows the relationship between the number of alleles and the gene identity, F, observed for 28 mitochondrial genes within five geographic regions. For illustration, we have also plotted the 95% confidence region obtained by computer simulation described below, based on the average sample size (N = 29). Most points lie within the 95% confidence region; however, several observations lie above the upper limit of genetic identity expected for the average sample size.

To further explore the fit of the infinite alleles model, we directly compared the empirical distributions of allele frequencies to the expected distributions generated by computer simulations, following Stewart's (39) procedure. This procedure makes use of the fact that the expected distribution of allele frequencies under the infinite alleles model is fully specified by the sample size and the observed number of alleles. For each locus and population, we simulated a distribution of allele frequencies and obtained the gene identity (F), the number of singletons (S) or uniquely represented alleles, and the frequency of the most common allele (C). The distributions of these statistics were tabulated over 1000 simulations and were used to evaluate the significance of the observed values of F, S, and C. We repeated this procedure for each locus using the total number of alleles observed in the pooled sample.

A plot of the observed vs. expected F values within geographic regions reveals that, for most loci, the observed values tend to be larger than those predicted from the infinite alleles model (Fig. 2), and the deviations tend to be more pronounced for the pooled sample (see below). Table 2

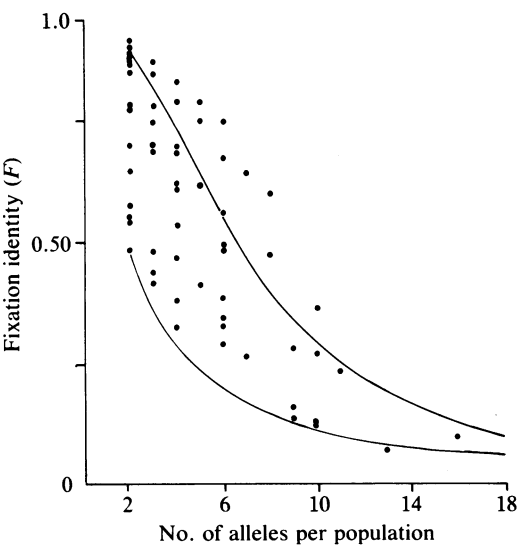

FIG. 1. Gene identity (F) plotted vs. the number of alleles per population for 28 mitochondrial loci sampled from five broad geographic regions. Solid lines represent upper and lower bounds of the 95% confidence region of F generated as described in the text.

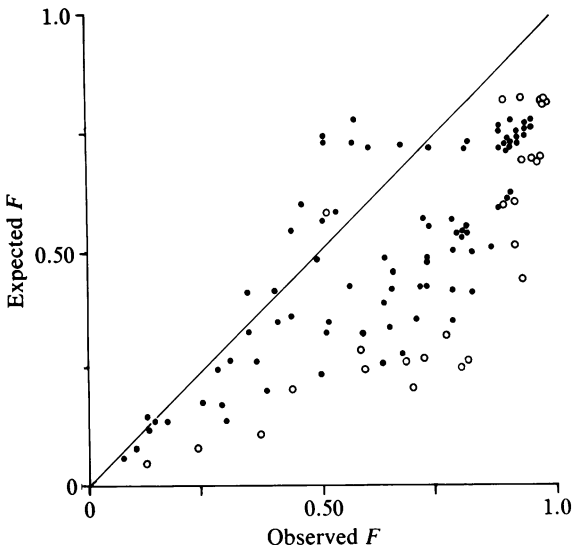

FIG. 2. Observed vs. expected gene identity for 28 mitochondrial loci within samples from geographic regions (●) and pooled sample (○). Expected gene identities were generated by computer simulation (see text).

summarizes the statistical comparisons to the neutral model for the different classes of mitochondrial genes. With the limited number of genomes sampled within geographic regions, only 35 of the 140 possible comparisons (28 loci × 5 populations) involved sufficiently large numbers of alleles for legitimate testing of the null hypothesis of neutrality. For example, simulations show that with five or fewer alleles in a sample of 20 genomes, even the most extreme F values are not statistically significant, because they occur in frequencies in excess of the significance level even under the neutral null hypothesis (see ref. 37). Of the 35 comparisons within geographic regions, 10 (29%) of the observed F values are significantly greater than the predicted values from the simulated distributions ($P < 0.025$). The greatest deviations from the neutral model are found for the protein-encoding loci (36% of 25 tests). The 10 tRNA loci are weakly polymorphic, averaging about 2.6 alleles per locus, and the single test that could be performed ($tRNA^{Val}$) showed a significant excess in F over the neutral expectation. In contrast, the noncoding sequence—i.e., D loop—showed no significant deviations within regions.

The observed excess in gene identity (F) means that certain loci have less allelic diversity than that predicted by the model of strictly neutral mutations. The basis for the lesser diversity is found by comparing the shapes of the observed and expected distributions of allele frequencies. In general, the observed distributions tend to be more leptokurtic than expected; that is, the most common alleles have higher frequencies than expected, alleles with intermediate frequencies are observed less often than expected, and singletons are found in greater numbers than expected under the neutral hypothesis. Common allele frequencies and the number of singletons observed within regions are given in Table 3 for the 12 most polymorphic loci. Common alleles have significantly greater frequencies than expected in 10 out of 35 tested cases (loci within regions), and singletons occur significantly more often than expected in 22 cases.

Departures from the neutral model are clearly evident in the pooled sample of 145 mtDNAs (Fig. 2). For 27 of 28 loci, F values in the pooled sample exceed the expected values (the exception is NADH dehydrogenase subunit 3). For 17 loci in which statistical comparisons were possible, the fraction of significant deviations of F increased in all groups of loci, except the tRNAs (Table 2). The genetic diversity at a locus tends to be lower in the pooled sample in comparison with the neutral predictions due to significant excesses in the frequency of common alleles (14 out of 17 loci) and in the numbers of singletons (13 out of 17 loci).

Table 2. Summary statistics and comparisons of observed genetic identities with the expected values from the infinite alleles model

Locus	Loci, no.	Alleles, average no.	Diversity, average	Proportion significant: Within regions	Proportion significant: Pooled sample
Noncoding	3	15.3	0.306	0 (5)	1.00 (2)
Proteins	13	11.8	0.306	0.36 (25)	0.92 (12)
rRNAs	2	10.5	0.318	0.25 (4)	1.00 (2)
tRNAs	10	2.6	0.050	1.00 (1)	0 (2)
All loci	28	8.8	0.215	0.29 (35)	0.82 (17)

For each class of loci, the proportion significant is the fraction of polymorphic loci for which the probability of obtaining a genetic identity exceeding the observed F was <0.025. The total number of tests performed is given in parentheses.

Table 3. Frequencies of the most common allele in each geographic region and the number of singletons (S) observed for 12 loci of the mtDNA genome with five or more alleles

Locus	Allele no.	Geographic region: I (20)	II (34)	III (21)	IV (44)	V (26)	Total (145)	χ^2
D loop	1	0.25	0.26	0.33	0.43	0.15	**0.30**	38.1*
	2	0.00	0.00	0.00	0.00	0.27	0.05	
	S	**10**	**11**	5	**7**	5	**28**	
ND 5	1	0.25	0.44	0.24	0.59	0.04	**0.36**	80.0*
	2	0.35	0.32	0.57	0.21	0.27	0.32	
	3	0.00	0.00	0.00	0.00	0.46	0.08	
	S	**6**	**8**	**4**	**7**	4	**20**	
CytOxase I	1	0.30	**0.71**	**0.81**	0.50	0.62	**0.59**	14.5*
	S	5	**5**	**4**	**4**	**3**	**16**	
Cyt *b*	1	0.75	0.65	0.52	**0.80**	0.38	**0.64**	90.5*
	2	0.00	0.00	0.00	0.00	0.62	0.11	
	S	NS	**2**	3	**5**	NS	**7**	
ND 4	1	**0.55**	0.85	0.95	**0.89**	0.85	**0.83**	14.8*
	S	**7**	**2**	NS	**5**	NS	**8**	
ND 1	1	0.40	0.77	0.81	**0.89**	0.81	**0.77**	41.8*
	2	0.45	0.03	0.00	0.00	0.00	0.07	
	S	NS	**4**	NS	**3**	NS	**6**	
CytOxase III	1	**0.90**	**0.82**	1.00	**0.93**	0.88	**0.90**	5.3
	S	NS	**6**	NS	**3**	NS	**7**	
CytOxase II	1	**0.90**	0.79	0.86	0.96	0.96	**0.90**	7.0
	S	NS	**3**	NS	NS	NS	**8**	
ATPase 6	1	0.85	0.79	0.91	0.70	0.96	**0.82**	8.8
	S	NS	1	NS	2	NS	**5**	
16S rRNA	1	0.55	0.85	0.86	**0.91**	0.96	**0.85**	17.7*
	S	NS	**2**	NS	**4**	NS	**7**	
12S rRNA	1	0.90	0.53	0.81	**0.89**	0.65	**0.75**	17.3*
	S	NS	2	NS	1	NS	**5**	
ND 2	1	0.95	**0.71**	1.00	**0.84**	1.00	**0.88**	17.2*
	S	NS	**3**	NS	**3**	NS	3	

The number of mtDNA genomes sampled from each area is given in parentheses. Boldface type designates entries in which the probability of the observed value exceeding the expected value under the hypothesis of strict neutrality is less than 0.025. NS indicates that the number of alleles was insufficient to perform the Ewens–Watterson test. Asterisks designate significant ($P < 0.05$) heterogeneity in common allele frequencies across regions as indicated by contingency χ^2 tests. ND, NADH dehydrogenase; Cyt, cytochrome; CytOxase, cytochrome oxidase.

Analysis of Population Subdivision. Differences among subpopulations in gene frequencies can be quantified by Nei's (31) coefficient of population differentiation, G_{ST}, defined as $(H_T - H_S)/H_T$, where H_T is the unbiased estimate of genetic diversity calculated from allele frequencies in the total population (pooled sample). For mtDNA genes, G_{ST} ranges from 0.147 for cytochrome *b* to 0 for several tRNAs and has an average of 0.063 over 28 loci (Table 1). These values suggest that gene flow among geographic regions has been sufficiently limited so that genetic differences among mtDNA gene pools have accumulated.

In principle, the degree of genetic differentiation among subpopulations, as measured by G_{ST}, can be attributed to both variation in the frequency of alleles found in all subpopulations and to the occurrence of alleles in a limited number of subpopulations. Heterogeneity tests for differences in common allele frequencies indicate significant geographic variation in allele frequencies for 10 of 12 highly polymorphic loci (Table 3). Because of the large number of singletons, most alleles are private alleles, observed only in a single geographic region. However, in Papua New Guinea, private alleles at three loci (D loop, NADH dehydrogenase subunit 5, and cytochrome *b*) have achieved relatively high frequencies (Table 3), suggesting that this population has experienced limited gene flow with other mtDNA gene pools (40).

Variation and Sequence Length. There is a positive relationship between the degree of allelic variation at a mtDNA locus and the size of the gene as measured by sequence length. Locus size (in base pairs) is significantly correlated with the number of polymorphic sites per locus ($r = 0.82$), the number of alleles per locus ($r = 0.70$), the average genetic diversity within regions ($r = 0.72$), and the relative differentiation as measured by G_{ST} ($r = 0.73$). These correlations remain significant when the tRNAs and smaller loci of less than 100 base pairs are omitted from the analysis and when G_{ST} values were corrected for sampling variance (41). Because of the ascertainment bias associated with restriction site data (16, 18, 42), rare sites can cause an underestimation of F and perhaps generate spurious correlations. When rare sites are disregarded, the correlation between G_{ST} and size becomes $r = 0.72$, indicating that the correlation is not generated by ascertainment bias.

Analysis of Site Gains. The exact position and nature of those base substitutions that account for the presence of a restriction site in a given mtDNA and its absence in a known sequence can usually be inferred by assuming only a single substitution per restriction site (12). Of the 90 restriction site gains in which the exact nucleotide substitution could be inferred, 61 occur in coding regions, and 21 of these result in an amino acid replacement. This proportion of replacement substitutions (0.34) is less than half the expected fraction (0.78) estimated from the mitochondrial genetic code and the codon frequencies in the Cambridge sequence (28). Such a low frequency of replacement substitution is compelling evidence that replacement substitutions have been selectively disfavored (43). These results further indicate that departures from the infinite alleles model are not all due to the past action of selection on a single non-neutral site.

DISCUSSION

The availability of high-resolution restriction maps of human mtDNA permits one to define alleles at mitochondrial loci based on patterns of restriction site polymorphism and thus to make use of statistical techniques developed for the study of the factors influencing allelic variation in natural populations (31, 35, 36, 39–47). In particular, the infinite alleles model provides theoretical predictions of the stationary distribution of allele frequencies that can be directly compared to observed distributions (44–46). The analysis of allelic variation at mitochondrial loci, using the Ewens–Watterson sampling theory, reveals that most of the genetic diversity within geographic regions falls in the range predicted by the infinite alleles model, from which we infer that neutral mutations have been a major influence in the evolutionary divergence of human mitochondrial genomes.

The distributions of allele frequencies contain some departures from the neutral mutation model: common alleles occur in higher frequency than expected, especially for loci encoding proteins and tRNAs, and singletons are more numerous than expected for all classes of loci. By systematically examining the assumptions of Ewens–Watterson sampling theory, one may attribute these departures to several sources. For example, the theory assumes complete identification of all allelic types (47). Despite the fact that restriction mapping can detect a greater proportion of existing sequence diversity than other indirect techniques (e.g., sequential electrophoresis, refs. 45 and 46), some variation within allelic classes can remain undetected. Unidentified alleles may also arise through convergence of different sequences to the same restriction site haplotype through repeated gains and losses of specific sites. However, the probability of convergence to an identical allelic state becomes extremely small for loci with a large number of polymorphic sites. Similarly, polymorphic sites will most likely delineate new alleles when the sample size and neutral mutation parameter ($N\mu$) are sufficiently large (48).

A second assumption underlying the statistical theory is that the distribution of allele frequencies has achieved a steady state between the mutational gain of new variants and the stochastic loss of alleles through genetic drift (36, 37). Because of the recent expansion of the human population, such a steady state has probably not yet been reached (49). Theory predicts that during the period of population expansion the expected number of neutral alleles increases more rapidly than the expected genetic diversity (50). This prediction agrees with the deviations of the observed distributions of alleles detected by the Ewens–Watterson tests, especially with the excess in singletons observed at nearly every locus in the mitochondrial genome.

A third assumption is that the population is geographically unstructured and homogeneous with respect to gene frequencies (47, 51). If migration is limited within regions, microgeographic variation in allele frequencies would create deviations from the neutral model in the same manner as detected in the pooled sample. As with the effect of population expansion on genetic variation, microgeographic structuring would affect all loci in the mitochondrial genome; however, at this point we have insufficient information to assess the extent of fine-scale structuring.

A final assumption of the Ewens–Watterson theory is that all alleles at a locus represent strictly neutral mutations whose dynamics have been affected solely by random genetic drift. As an alternative to the effect of population expansion, the direction of departure of the observed allele frequencies from the infinite alleles model may be attributed to the past action of purifying selection favoring the common allele within each population. The contribution of purifying selection is further supported by the observation that the probability that an inferred replacement site is a rare site, a site that occurs only once in the sample, is significantly greater than expected by chance alone.

The complete linkage of mitochondrial genes, resulting from the apparent lack of a recombination mechanism for mtDNA molecules, has two important implications in interpreting the results of the tests of neutrality. First, the random sampling of mtDNA genomes over evolutionary time can generate positive correlations in genetic diversities for completely linked loci, even if all mutations are strictly neutral

(29, 52). However, computer simulations indicate that these correlations are insufficient to inflate the probability of deviations of genetic diversities in the Ewens–Watterson test (A.G.C. and T.S.W., unpublished data). Second, in the absence of genetic recombination, neutral and nearly neutral alleles may increase in frequency through their association with a selectively favored mutation that has occurred in any part of the genome (53–55). Adaptive mutations can increase the effect of genetic drift by reducing the effective population size, thus reducing the average genetic diversity (54, 55). These indirect effects of selection on linked neutral alleles would be difficult to attribute to any given locus, but computer simulations (A.G.C. and T.S.W., unpublished data) show that even with strong selection a linked neutral locus may still adequately fit the allele frequency spectrum predicted by the Ewens–Watterson theory. Our argument that more than one site has faced selection is based on the excess of silent substitutions seen throughout the genome.

Our analyses indicate that several evolutionary factors are responsible for the levels of variation of mitochondrial genes in the human population. We infer that neutral mutations and genetic drift have strongly influenced the levels of genetic variation at mitochondrial loci, because (*i*) many observations fall within the range of values predicted by the infinite alleles model, (*ii*) noncoding sequences show the smallest departures from the neutral model, (*iii*) most restriction site changes in coding regions involve silent substitutions, and (*iv*) genetic diversity correlates with sequence length. The fact that genetic drift can strongly affect organelle genes through the reduction in effective population size resulting from maternal transmission is well known (21, 25). In addition, it is likely that the recent expansion of the human population and some form of selection have distorted the distributions of allele frequencies away from the steady-state distributions expected for strictly neutral mutations.

The method of constructing allelic identities from restriction patterns can be applied to any genetic region where the correspondence between the genetic map (location of structural genes) and the restriction map is precisely known. In particular, the analysis we employ will be valuable in cases where the nucleotide sequence is known and where large population samples of restriction site variation can be obtained. The results of these analyses will allow a new level of resolution of molecular variation at a fraction of the effort of complete sequencing.

We thank Drs. Thomas and Barbara Ryan, directors of Minitab, Inc., for generous donation of computer time and two reviewers for comments on the manuscript. This work was supported by grants from the National Institutes of Health and the National Science Foundation.

1. Ferris, S. D., Wilson, A. C. & Brown, W. M. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2432–2436.
2. Brown, G. G. & Simpson, M. V. (1981) *Genetics* **97,** 125–143.
3. Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. (1982) *J. Mol. Evol.* **18,** 225–239.
4. Lansman, R. A., Avise, J. C., Aquadro, C. F., Shapira, J. F. & Daniel, S. W. (1983) *Evolution* **37,** 1–16.
5. Cann, R. L. & Wilson, A. C. (1983) *Genetics* **104,** 699–711.
6. Powell, J. R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 492–495.
7. George, M., Puentes, L. A. & Ryder, O. A. (1983) in *International Stud Book of African Rhinoceroses*, ed. Klös, H. G. (Zöologischer Gärten, Berlin), No. 2, pp. 60–67.
8. Greenberg, B. D., Newbold, J. E. & Sugino, A. (1983) *Gene* **21,** 33–49.
9. Ferris, S. D., Sage, R. D., Prager, E. M., Ritte, U. & Wilson, A. C. (1983) *Genetics* **105,** 681–721.
10. Olivo, P. D., Van de Walle, M. J., Laipis, P. J. & Hauswirth, W. W. (1983) *Nature (London)* **306,** 400–402.
11. Avise, J. C. & Saunders, N. C. (1984) *Genetics* **108,** 237–255.
12. Cann, R. L., Brown, W. M. & Wilson, A. C. (1984) *Genetics* **106,** 479–499.
13. Kessler, L. G. & Avise, J. C. (1985) *Mol. Biol. Evol.* **2,** 109–125.
14. Wilson, A. C., Cann, R. L., Carr, S. M., George, M., Gyllensten, U. B., Helm-Bychowski, K. M., Higuchi, R. G., Palumbi, S. R., Prager, E. M., Sage, R. D. & Stoneking, M. (1985) *Biol. J. Linn. Soc.* **26,** 375–400.
15. Nei, M. & Li, W.-H. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5269–5273.
16. Engels, W. R. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 6329–6333.
17. Nei, M. & Tajima, F. (1981) *Genetics* **97,** 145–163.
18. Ewens, W. J., Spielman, R. S. & Harris, H. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3748–3750.
19. Takahata, N. & Maruyama, T. (1981) *Genet. Res.* **37,** 291–302.
20. Chapman, R. W., Stephens, J. C., Lansman, R. A. & Avise, J. C. (1982) *Genet. Res.* **40,** 41–57.
21. Birky, C. W., Maruyama, T. & Fuerst, P. (1983) *Genetics* **103,** 513–527.
22. Takahata, N. & Slatkin, M. (1983) *Genet. Res.* **42,** 257–265.
23. Aquadro, C. F., Kaplan, N. & Risko, K. J. (1984) *Mol. Biol. Evol.* **1,** 423–434.
24. Clark, A. G. (1984) *Genetics* **107,** 679–701.
25. Avise, J. C., Neigel, J. E. & Arnold, J. (1984) *J. Mol. Evol.* **20,** 99–105.
26. Takahata, N. & Palumbi, S. R. (1985) *Genetics* **109,** 441–457.
27. Stoneking, M., Bhatia, K. & Wilson, A. C. (1986) in *Genetic Variation in Human Populations in the Tropics*, eds. Roberts, D. F. & DeStefano, G. F. (Cambridge Univ. Press, Cambridge, U.K.), in press.
28. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1981) *Nature (London)* **290,** 457–465.
29. Tajima, F. (1983) *Genetics* **105,** 437–460.
30. Cann, R. L., Stoneking, M. & Wilson, A. C. (1986) *Nature (London)*, in press.
31. Nei, M. (1975) *Molecular Population Genetics and Evolution* (Elsevier, New York).
32. Nei, M. & Roychoudhury, A. K. (1974) *Genetics* **76,** 379–390.
33. Nei, M. (1978) *Genetics* **89,** 583–590.
34. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, New York).
35. Ewens, W. J. (1972) *Theor. Popul. Biol.* **3,** 87–112.
36. Ewens, W. J. (1979) *Mathematical Population Genetics* (Springer, New York).
37. Watterson, G. A. (1978) *Genetics* **88,** 405–417.
38. Kimura, M. & Crow, J. F. (1964) *Genetics* **49,** 725–738.
39. Fuerst, P. A., Chakraborty, R. & Nei, M. (1977) *Genetics* **86,** 455–483.
40. Slatkin, M. (1985) *Evolution* **38,** 53–65.
41. Wright, S. (1978) *Evolution and the Genetics of Populations* (Univ. of Chicago Press, Chicago).
42. Hudson, R. R. (1982) *Genetics* **100,** 711–719.
43. Kreitman, M. (1983) *Nature (London)* **304,** 412–417.
44. Hedrick, P. W. & Thomson, G. (1983) *Genetics* **104,** 449–456.
45. Keith, T. P. (1983) *Genetics* **105,** 135–155.
46. Keith, T. P., Brooks, L. D., Lewontin, R. C., Martinez-Cruzado, J. C. & Rigby, D. L. (1985) *Mol. Biol. Evol.* **2,** 206–216.
47. Ewens, W. J. & Gillespie, J. H. (1974) *Theor. Popul. Biol.* **6,** 35–57.
48. Griffiths, R. C. (1982) *Adv. Appl. Probab.* **14,** 225–239.
49. Nei, M. & Graur, D. (1984) *Evol. Biol.* **17,** 73–118.
50. Nei, M. & Li, W.-H. (1976) *Genet. Res.* **28,** 205–214.
51. Slatkin, M. (1982) *Genetics* **100,** 533–545.
52. Hill, W. G. (1975) *Theor. Popul. Biol.* **8,** 117–126.
53. Thomson, G. (1977) *Genetics* **85,** 753–788.
54. Hedrick, P. W. (1982) *Bioscience* **32,** 845–853.
55. Levin, B. R. (1981) *Genetics* **99,** 1–23.