Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 839–843, February 1991
Evolution

Drift, admixture, and selection in human evolution: A study with DNA polymorphisms

(population genetics/simulation/neutral theory)

ANNE M. BOWCOCK*†, JUDITH R. KIDD‡, JOANNA L. MOUNTAIN*, JOAN M. HEBERT*, LUCIANO CAROTENUTO§, KENNETH K. KIDD‡, AND L. LUCA CAVALLI-SFORZA*

*Department of Genetics, Stanford University, Stanford, CA 94305; ‡Department of Human Genetics, Yale University, New Haven, CT 06510; and §Dipartimento di Sistemi, Universita della Calabria, 87036 Arcavacata, Cosenza, Italy

Contributed by L. Luca Cavalli-Sforza, October 26, 1990

ABSTRACT **Accuracy of evolutionary analysis of populations within a species requires the testing of a large number of genetic polymorphisms belonging to many loci. We report here a reconstruction of human differentiation based on 100 DNA polymorphisms tested in five populations from four continents. The results agree with earlier conclusions based on other classes of genetic markers but reveal that Europeans do not fit a simple model of independently evolving populations with equal evolutionary rates. Evolutionary models involving early admixture are compatible with the data. Taking one such model into account, we examined through simulation whether random genetic drift alone might explain the variation among gene frequencies across populations and genes. A measure of variation among populations was calculated for each polymorphism, and its distribution for the 100 polymorphisms was compared with that expected for a drift-only hypothesis. At least two-thirds of the polymorphisms appear to be selectively neutral, but there are significant deviations at the two ends of the observed distribution of the measure of variation: a slight excess of polymorphisms with low variation and a greater excess with high variation. This indicates that a few DNA polymorphisms are affected by natural selection, rarely heterotic, and more often disruptive, while most are selectively neutral.**

This paper presents results of the first phase of our study of human evolution based on nuclear DNA variation in five populations studied for a selected set of 100 DNA polymorphisms. At this stage we have chosen to focus on a large number of markers typed on each of a small number of defined populations because the reliability of conclusions is so dependent on the number of independent markers (1–3). Using these data (presented in detail in ref. 38), we examine the shape of the tree and compare it with earlier trees, investigate an observed anomaly of evolutionary rates, and study variation of different genes and its bearing on the theory of neutral evolution.

Prior studies of human populations using DNA markers have been limited almost entirely to mitochondrial DNA (4, 5) and the much less informative β-globin region (6) or have involved combinations of data from different populations for a great variety of loci (7). These studies supported previous conclusions (8) regarding the early separation of African and non-African lineages, based on classical polymorphisms (studied by electrophoresis of proteins and immunological tests). A further, more extensive analysis of classical markers has confirmed this conclusion (9).

There are now >2000 (2064 as of July 1990; unpublished observation) catalogued and documented DNA polymorphisms. Most of these are standard restriction fragment length polymorphisms (RFLPs) (10–12), but some of the more recently discovered polymorphisms have been typed using the polymerase chain reaction (13, 14). The two techniques provide an enormous, previously untapped source of genetic markers. As discussed elsewhere (38), the data set analyzed here is the beginning of a reference set for future comparisons of a large number of populations.

Data

The populations typed for 100 DNA markers were (*i*) individuals of European origin from ongoing studies in our laboratories or reported in the literature; (*ii*) Chinese born in mainland China living in the San Francisco Bay Area; (*iii*) non-Austronesian speaking Melanesians from Bougainville (blood samples collected by J. Friedlaender); and (*iv* and *v*) two groups of African Pygmies, one from the Central African Republic and the other from northeastern Zaire. Thus they are fairly representative of the world's aboriginal populations, except for Amerind populations, who are presently being typed. The Pygmies of the Central African Republic are probably an admixture of about three-fourths non-Pygmy African ancestry and one-fourth Pygmy ancestry of the Zaire type (15) on the basis of classical markers; this is reflected in their taller size compared with the other Pygmy group, from Zaire, believed to be less mixed and more representative of these hunter–gatherer populations.

In order to obtain sufficient DNA, B lymphocytes from samples of the selected indigenous populations were transformed with Epstein–Barr virus. DNA from these cell lines was tested for polymorphism of both known and anonymous DNA segments (16, 17). The average number of chromosomes screened per population is 50.3 for African Pygmies from Zaire, 43.7 for Pygmies from the Central African Republic, 109.7 for Europeans, 63.8 for Chinese, and 25.8 for Melanesians. We previously reported (17) detailed data from the same five populations including methods and probes used for 47 RFLPs. Data for the remaining polymorphisms are published separately (38). Statistical reliability and agreement with evolutionary models were tested by bootstrap resampling of the polymorphisms (9, 18, 19).

The Evolutionary Tree

Observed genetic distances are shown in Table 1. The tree calculated by maximum likelihood (23, 24) for a model assuming constant evolutionary rates is shown in Fig. 1*a*. This tree confirms the hypothesis that the earliest divergence in human evolution separated Africans and non-Africans (5,

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.

Abbreviations: RFLP, restriction fragment length polymorphism; kyr, kiloyear(s).

†Present address: Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX 75235.

Table 1. Estimates of genetic distances among five populations, based on data for 99 DNA polymorphisms

	PygC	PygZ	Eur	Chi	Mel
PygC	—	0.023	0.088	0.144	0.133
PygZ	0.043	—	0.084	0.139	0.139
Eur	0.141*†	0.142*†	—	0.058	0.086
Chi	0.235*‡	0.235*‡	0.093	—	0.094
Mel	0.242*‡	0.265*‡	0.148	0.171	—

Distances above the diagonal are Nei's unbiased estimates of standard genetic distance (20). Those below the diagonal are based on F_{ST} (21), also corrected for sampling error due to the small number of individuals tested. The two distances are proportional to each other in the observed range, with a ratio of the F_{ST} distance to Nei's standard genetic distance of 1.80 ± 0.04. The five populations are Pygmies from the Central African Republic (PygC), Pygmies from Zaire (PygZ), Caucasoids of European origin (Eur), Chinese (Chi), and Melanesians (Mel). According to a treeness test (22), given the tree of Fig. 1*a* the six distances between the two African populations and the three non-African populations (*) are expected to be equal. This expectation under constant evolutionary rates is not borne out by the data. A significance test by bootstrap of the difference between the average of the African–European distances (†) and the average of the African–Chinese and African–Melanesian distances (‡) ($t = -0.1028 \pm 0.0252$, $P < 0.0001$) shows that Africans are significantly closer genetically to Europeans than to Chinese or Melanesians. This probability is sufficiently small to compensate for the fact that we have chosen one of several possible comparisons of the six distances. For the F_{ST} simulation an additional polymorphism was included; the differences in genetic distance are minor.

6, 8, 9). The second fission suggested by these data separates Melanesians from Chinese plus Europeans. This is consistent with the observation based on classical genetic markers that the second divergence separated Australians, New Guineans, and Southeast Asians from North Eurasians (and Amerinds, not tested here) (9).

The fit of this tree, however, is not acceptable. If genetic evolution in different lineages is independent and occurs at equal rates, we expect all distances between pairs of populations generated by a specific tree node to be equal (22). For example, all distances between African and non-African populations are expected to be equal. It is clear from Table 1 that the observed distances do not agree with this expectation: the distances between the African populations and Europeans are significantly smaller than the distances between the African populations and the other non-African populations.

With alternative approaches to tree construction that allow evolutionary rates to vary, the branch corresponding to Europeans is very short, and shorter than the Chinese branch (Fig. 2). Because of the lack of specific hypotheses on evolutionary rates, the trees are all unrooted and their topology is slightly different from that of Fig. 1*a*, with the European branch splitting off before the separation of Chinese and Melanesians. Short lengths for European and related branches have been observed previously for trees constructed from classical genetic data (1, 8, 30). Two possible explanations for the short European branch are (*a*) that after the fission, Europeans diverged at a much lower evolutionary rate, or (*b*) that Europeans are descendants of a population that arose due to admixture between two ancestral populations. The ad hoc hypothesis of a lower evolutionary rate in Europe is not further testable. We can, however, rule out one possible cause of such a reduction of evolutionary rates: the increase in population density due to agriculture. This was of such magnitude (31) that it may have frozen genetic drift in Europe. However, because this increase in density occurred fairly recently relative to the time of settlement of modern humans in Europe, it cannot have caused a reduction of more than 20–25% in the evolutionary rate of Europeans; trees such as that of Fig. 2 indicate a reduction of the order of 80–90%. In contrast to the lower

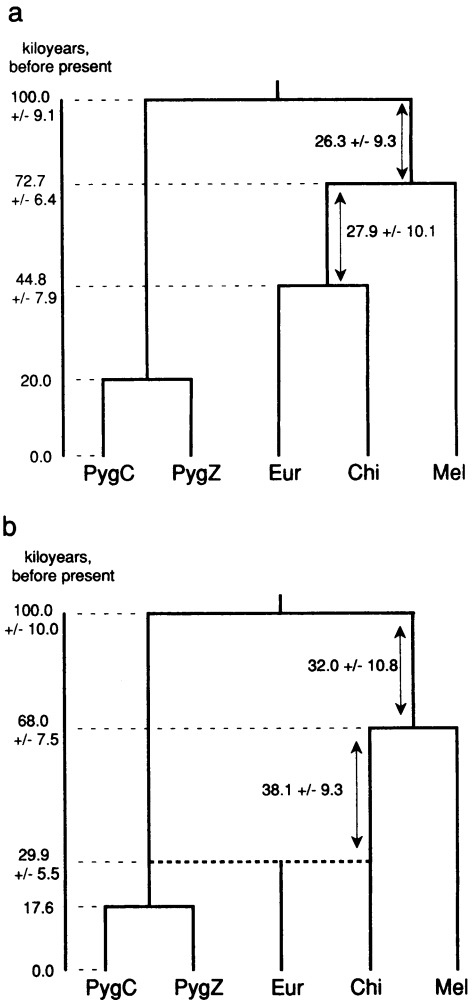
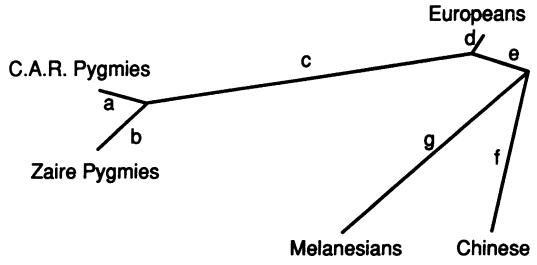


FIG. 1. (*a*) Rooted tree constructed by maximum likelihood, assuming constant evolutionary rates, for the five populations whose distances are shown in the lower triangle of Table 1. Eighty-four percent of bootstrap replicates give the tree topology of *a*, supporting the sequence of fissions. Standard errors of branch lengths were estimated using the bootstrap. The time scale was based on the hypothesis that the separation of Africans from non-Africans took place 100,000 years ago (25). (*b*) Tree constructed by maximum likelihood, assuming a model of admixture between ancestral Africans and ancestral Asians, fitting the distances of the lower triangle of Table 1. According to this model two divergent populations contribute in specified proportions to form a new population. Various pairs of ancestral populations from which the European branch may have descended by admixture were tested for choosing ancestral types that contributed to the admixture. Data were found to be most consistent with this tree; ancestral Europeans are estimated to be an admixture of 65% ancestral Chinese and 35% ancestral Africans.

evolutionary rate hypothesis, the hypothesis that the shorter branch leading to Europeans is due to admixture is testable. This hypothesis was suggested earlier for the analysis of three populations (Africans, Europeans, and Northeast Asians; ref. 7) but was not quantitatively analyzed. One can show that a branch to a population resulting from admixture tends to be shorter than other branches when methods are used that do not require constant evolutionary rates, by a simple extension of the theoretical treatment of admixture between branches of a tree (22).

Europeans as an Early Admixture

We used a modified maximum-likelihood method to test the general model of admixture, using formulas derived from



Method	Branch a	b	c	d	e	f	g
minimum string*	13.5	15.2	56.1	11.3	7.7	41.1	74.6
neighbor-joining	17.5	25.5	111.1	8.9	26.1	67.7	103.9
least squares*	19.4	23.6	113.6	6.9	24.4	65.3	107.7

*assuming neighbor-joining topology

FIG. 2. Tree constructed using Nei's neighbor-joining method, from genetic distances of the lower triangle of Table 1. The branch lengths estimated by three different methods, including "neighbor-joining" (20), "minimum-string," and "least-squares" (1, 26–29), are given. The European branch joins the tree between the branch leading to Africans and that leading to Chinese plus Melanesians; branch e is significantly different in length from zero, according to bootstraps of the minimum-string tree for the neighbor-joining topology. C.A.R., Central African Republic.

those given in ref. 22. The data were found to be consistent with admixture between the branch leading to Chinese after their separation from Melanesians and the branch leading to the two African populations (Fig. 1*b*). From maximum likelihood estimates the European admixture consisted of 65% Chinese ancestors and 35% African ancestors (with a standard error of ±8%) and took place at a time ≈70% of the total since the origin, or 30 ± 6 kiloyears (kyr) ago. Based on the admixture model, the other divergence-time estimates are also modified (Fig. 1*b*). In particular, the time of separation of Melanesians and Chinese is slightly decreased (to 68 ± 8 kyr) compared with the estimate based on the analysis without admixture.

This admixture model fits the data very well. Its major weakness is that it is extremely specific in assuming that admixture took place rapidly, followed immediately by separation. While this series of events is possible, it seems unrealistic. Further modification of the model to incorporate gene flow between the ancestral African and ancestral Asian neighboring populations over a span of several thousand years would probably change the fit very little. Incorporation of cycles of admixture would have a similar effect and might also be more realistic. One might even postulate a continuous admixture, in time, in space, or in both: a chain of populations somewhat similar to a stepping-stone model in which the ancestors of Europeans are geographically intermediate between the two extremes, Africans and Asians. Similar models, simulated and tested with tree programs, give rise to trees having a comb structure with short intermediate teeth (30), not unlike the observed tree. The actual situation was probably somewhere between the two extremes of instantaneous admixture and continuous flow; it is unlikely that quantitative analysis of these data will allow us to choose satisfactorily from among these two extremes and the intermediate situations. The intermediacy of the Europeans with respect to the African and East Asian populations may also be due in part to bias in the selection of polymorphisms; most of them were selected because they are polymorphic in European populations. The average heterozygosity for these markers in the European population (0.373 ± 0.014), however, is not much higher than in the other four populations (Chinese, 0.340 ± 0.018; Central African Republic Pygmies, 0.313 ± 0.019; Zaire Pygmies, 0.304 ± 0.019; Melanesians, 0.267 ± 0.020). It does not seem likely that the bias thus generated could be responsible for the observed intermediacy of Europeans. Testing this hypothesis requires using polymorphisms identified in other populations.

Historical evidence suggests that admixture was likely. The expansion of modern humans who replaced Neanderthals in Europe probably originated in West Asia; the languages spoken in paleolithic Europe, of which Basque is the only widely accepted surviving descendant, appear to have some residual similarity to languages from the Caucasus in West Asia (32). The population that originally replaced Neanderthals 30–40 kyr ago may have already been a mixture between Africans and East Asians, because West Asia is geographically intermediate between Africa and East Asia. A later expansion from the Near East toward Europe, Asia, and North Africa began with the spread of farmers 9–10 kyr ago (31). The estimated date of this last event is later than expected based on the 30-kyr date calculated for the single admixture, but this may have been one of several cycles of admixture.

The advantage of the admixture model we have suggested is that it fits the genetic data very well; it is likely that many other admixture models would do the same, but those models are more complex and less easy to study. We therefore choose to retain our simple admixture model and are prepared to replace it with other models giving good formal fits, if or when additional data from other populations or new external evidence give them preferential support. The model is also relevant for our next aim, the test of natural selection.

Natural Selection or Neutral Variation?

The 100 polymorphisms differ considerably in variation of gene frequencies across populations. The most stable one (*HP/Bam*HI) has the lowest gene frequency (0.35) in Chinese and the highest (0.46) in Zaire Pygmies. The most variable one (*ADH2/Rsa* I) has the lowest frequency (0.13) in Chinese and the highest (1.00) in the two Pygmy populations. Data are published in detail elsewhere (38). The usual measure of variation of a polymorphism among populations, F_{ST},[¶] has a theoretical range of 0–1; the two polymorphisms mentioned above have F_{ST} values of 0.00 and 0.49, respectively. The average F_{ST} value from all 100 polymorphisms is 0.139 ± 0.010. It does not differ significantly from that of the 120 "classical" (non-DNA) polymorphisms (0.119 ± 0.010) calculated for 42 world populations (used in ref. 9). The observed average F_{ST} value for the 61 known genes (0.144) does not differ significantly from that of the 39 anonymous segments (0.133). Nor does the average for the biallelic genes (0.139) differ from that of the multiallelic polymorphisms (0.140).

The variation of F_{ST} from one polymorphism to another may help us establish whether natural selection is playing a role or whether variation is selectively neutral. In the latter case the only force at play is drift; we expect this pressure to be equal for all genes, since drift depends only on demographic properties of the populations and not on the particular gene being studied (33). F_{ST} will still vary from gene to gene, but the extent of variation will be predictable. If natural selection is at play, and if in some environments one allele is favored while in others a different allele is at an advantage (disruptive selection), then F_{ST} for that gene is expected to be higher than for a gene affected by drift alone. If, on the other hand, natural selection favors the heterozygote over both

[¶] F_{ST} is the ratio of the variance of m gene frequencies, $\Sigma(p_i - \bar{p})^2/(m - 1)$, to the maximum value of this variance, $\bar{p}(1 - \bar{p})$, where p_i are the m gene frequencies and $\bar{p}$ is the average gene frequency.

homozygotes (stabilizing or balanced selection), then we can expect F_{ST} to be lower than average.

In order to evaluate the variation of F_{ST} in the five populations, we must first predict the distribution of observed F_{ST} values under drift alone. We can then compare this to the observed distribution to see whether there is evidence that some genes have significantly low or high F_{ST} values. An attempt to make this prediction for classical human polymorphisms (34) was inconclusive, partly because too few polymorphisms were used. In addition, the attempt was criticized (35) because the populations were assumed to have evolved independently from each other. In fact the evolutionary history of these populations is known at least approximately and produces constraints that affect the theoretical F_{ST} distribution under drift alone, as shown by Robertson (35).

In the present case the number of polymorphisms is large, few have reached fixation, and evolutionary history is reasonably well known. It was thus possible to set up a simulation that took evolutionary history into account. The conditions of the simulation are briefly summarized in Fig. 3, which shows two simulated distributions of F_{ST} for given initial gene frequencies, on which the distribution depends. The main assumptions of the simulation are that the average F_{ST} equals the observed one (0.139), gene frequencies are distributed as Beta variates (22, 34), and the population size in each branch remains constant. Simulations were carried out separately for the two evolutionary histories suggested by the two trees of Fig. 1. The second, more conservative case is represented in Figs. 3 and 4. It involves a modest decrease in the variation of simulated F_{ST} values compared with the first.

From the percentiles of distributions such as those of Fig. 3, for each of a number of initial gene frequencies from 0.01 to 0.99, we built the curves of Fig. 4. Percentiles (from 1% to 99%) of the F_{ST} distributions are plotted as a function of the initial gene frequency. Because the curves joining equal percentiles are symmetric around the gene frequency of 50%, only the gene frequency values between 0% and 50% are represented.

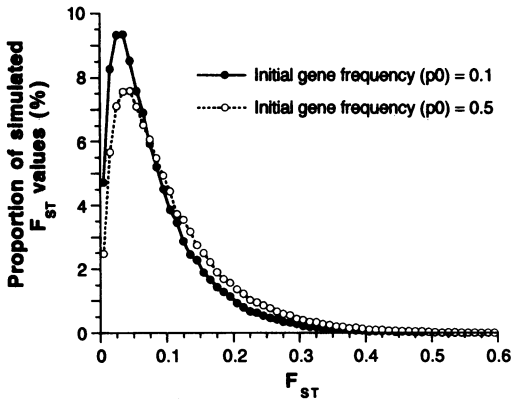


FIG. 3. Simulation of the F_{ST} distribution of five populations expected to average 0.139 at the end of evolution. The tree simulated was that of Fig. 1*b*. For each evolutionary interval (i.e., every segment of the tree) the gene frequency change was simulated by taking a random sample from a Beta distribution. Parameters of the distribution were based on (*a*) the initial gene frequency, equal to p_0 for segments starting at the origin and determined by the simulated process for lower segments, and (*b*) the fraction of the total F_{ST} (0.139) that is proportional to the relative length of the segment being simulated. For each initial gene frequency between 0.01 and 0.99 (steps of 0.01 between 0.01 and 0.10 and between 0.90 and 0.99; steps of 0.05 between 0.10 and 0.90), 100,000 F_{ST} values were calculated in order to generate distributions for each initial gene frequency, p_0. Two examples of such distributions are shown here: one with initial gene frequency p_0 = 10%, and one with initial gene frequency p_0 = 50%.

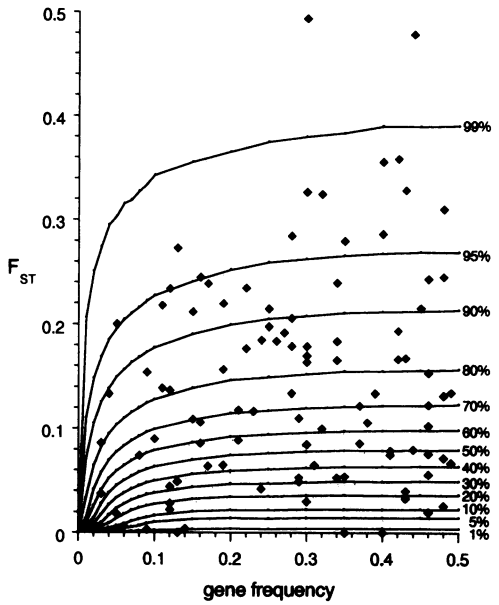


FIG. 4. F_{ST} values plotted versus the initial gene frequencies that were used to simulate the F_{ST} distributions, two of which are shown in Fig. 3. Curves represent the various percentiles of the simulated F_{ST} distributions for different initial gene frequencies. As these curves are symmetric around initial gene frequency of 50%, only the left half of the 0–100% range of initial gene frequencies is represented. Diamonds are the observed F_{ST} values for the 100 polymorphisms plotted against the corresponding mean gene frequency. When there are more than two alleles, the most common allele is represented. All average gene frequencies, x, greater than 50% are represented as $100 - x$. Two polymorphisms are represented by the diamond at $F_{ST} = 0$, gene frequency = 0.4.

The diamonds in Fig. 4 are the observed F_{ST} values of the 100 polymorphisms. If drift is the only force causing divergence between populations, one expects 1% of all observed F_{ST} values to be below the curve corresponding to the first (1%) percentile, 4% to be between the first percentile and the fifth percentile, and so on. Fig. 5 compares the number of F_{ST} values expected in the various percentile classes with the number observed. There is an excess of observed F_{ST} values at each end of the distribution, and therefore there are fewer in the middle. In other words, there are too many F_{ST} values that are small, and too many that are large. The overall difference between the numbers of expected and observed F_{ST} values is significant ($P = 0.0023$, Smirnov's χ^2 approximation).

Is this discrepancy due to natural selection? Before we accept such a strong statement we must consider a possible complication. In the drift model, population size is assumed to be constant. It is not easy to predict the effect that a deviation from this assumption would have on the distribution of F_{ST} values. It could cause an increase of variance, which might explain an excess of both low and high F_{ST} values. However, the effect would clearly depend on the extent of variation of the intensity of drift over time (in practice, on the extent of the variation of population size over time). We have no information regarding this point. One can, however, calculate a population size (average of all branches) that would lead to the observed average value of F_{ST}. The result is an effective population size of N_e = 13,000, which corresponds to an approximate total census size of 40,000 in every branch.‖ This is not an unrealistic estimate for popu-

‖The relationship between F_{ST}, time, and population size is: $F_{ST} = 1 - e^{-t/2N_e}$, where t is the number of generations of 25 years and N_e is the number of reproducing individuals, or effective population size, in each branch (36), which is roughly one-third of the census size (37).

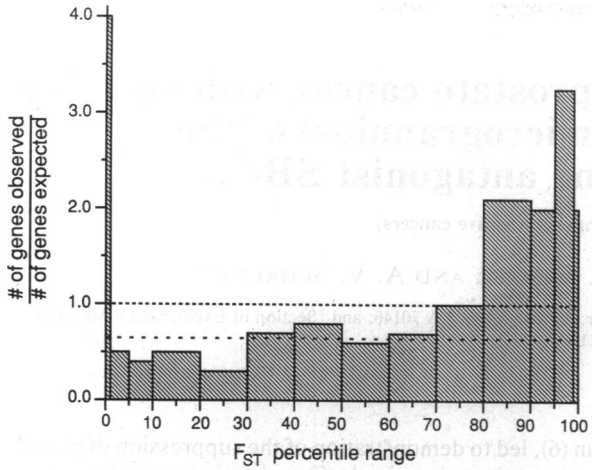


FIG. 5. Histogram showing the ratio of the number of genes observed to the number expected within each percentile range of the distribution of F_{ST} given in Fig. 4. The upper, heavy broken line indicates the expected rectangular distribution. If all F_{ST} values fit a model for neutral genes, their distribution would correspond to the horizontal line going through ordinate = 1. Deviation from this rectangular distribution is highly significant ($\chi^2 = 61.05$, 13 df, $P < 0.001$). Because an excess of F_{ST} values are low or high, the horizontal line corresponding to the fraction of genes that can be assumed to be neutral is below 1. The lower horizontal line corresponds to the best estimate of the distribution of neutral alleles, assuming that neutral alleles would give a rectangular distribution. Its height (0.63) indicates the minimum fraction of genes that behave as selectively neutral (assuming that population sizes are constant). The height of this lower broken line is calculated by progressively excluding classes of F_{ST} values at the left and right extremes on the abscissa, until the numbers of genes shown in the ordinate in the residual F_{ST} classes are not significantly different from equality as tested by χ^2. The χ^2 value corresponding to the lower line is 5.30 (8 df).

lation sizes of the main branches during the Paleolithic.

If the variation of population size over time does not contribute to the excess variation of F_{ST} between genes above that expected under drift alone, then the excess must be due to stabilizing natural selection at the lower end and disruptive selection at the upper end of F_{ST} values. One can estimate a minimum fraction of genes evolving under drift alone, and therefore selectively neutral, by calculating the fraction of polymorphisms compatible with a rectangular distribution of F_{ST} values. This is given by the lower broken horizontal line in Fig. 5 and corresponds to 63% of all polymorphisms.

A model of extreme fluctuation in population sizes might account for the observed F_{ST} variation, but it seems more likely that at least part of the deviation from the simulated distribution is due to the existence of either disruptive or stabilizing selection for a number of these genes. We estimate that the fraction of selected genes in our present data set may be as high as one-third, but this is likely to be a maximum estimate. If the variation of population size in time and space or other related factors have had a major effect in increasing the variation of F_{ST} values over and above that calculated here, then the estimate that two-thirds of the genes are selectively neutral is a conservative one.

We wish to thank J. Rogers for many helpful discussions. This research was supported by grants from the National Institutes of Health to L.L.C.-S. (GM20467) and to K.K.K. (MH39239 and GM40883) and from the National Science Foundation to K.K.K. (BNS8813234).

1. Cavalli-Sforza, L. L. & Edwards, A. W. F. (1964) in *Genetics Today, Proceedings of the 11th International Congress of Genetics*, ed. Geerts, S. J., (Pergamon, New York), Vol. 2, pp. 923–933.
2. Astolfi, P., Kidd, K. K. & Cavalli-Sforza, L. L. (1981) *Syst. Zool.* **30**, 156–169.
3. Pamilo, P. & Nei, M. (1988) *Mol. Biol. Evol.* **5**, 568–583.
4. Johnson, M. J., Wallace, D. C., Ferris, S. D., Rattazzi, M. C. & Cavalli-Sforza, L. L. (1983) *J. Mol. Evol.* **19**, 255–271.
5. Cann, R. L., Stoneking, M. & Wilson, A. C. (1987) *Nature (London)* **325**, 31–36.
6. Wainscoat, J. S., Hill, A. V. S., Boyce, A. L., Flint, J., Hernandez, M., Thein, S. L., Old, J. M., Lynch, J. R., Falusi, A. G., Weatherall, D. J. & Clegg, J. B. (1986) *Nature (London)* **319**, 491–493.
7. Nei, M. & Livshits, G. (1989) *Hum. Hered.* **39**, 276–281.
8. Nei, M. (1978) *Jpn. J. Hum. Genet.* **23**, 341–369.
9. Cavalli-Sforza, L. L., Piazza, A., Menozzi, P. & Mountain, J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6002–6006.
10. Kan, Y. W. & Dozy, A. M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5631–5635.
11. Botstein, D., White, R., Skolnik, M. & Davis, R. W. (1980) *Am. J. Hum. Genet.* **32**, 314–331.
12. Wyman, A. R. & White, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6754–6758.
13. Mullis, K. & Faloona, F. (1987) *Methods Enzymol.* **155**, 335–350.
14. Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1986) *Nature (London)* **324**, 163–166.
15. Cavalli-Sforza, L. L. (1986) *African Pygmies* (Academic, Orlando, FL).
16. Cavalli-Sforza, L. L., Kidd, J. R., Kidd, K. K., Bucci, C., Bowcock, A. M., Hewlett, B. S. & Friedlaender, J. S. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **51**, 411–417.
17. Bowcock, A. M., Bucci, C., Hebert, J. M., Kidd, J. R., Kidd, K. K., Friedlaender, J. S. & Cavalli-Sforza, L. L. (1987) *Gene Geography* **1**, 47–64.
18. Efron, B. (1982) *The Jacknife, Bootstrap, and Other Resampling Plans* (Soc. for Industrial and Appl. Math., Philadelphia).
19. Felsenstein, J. (1985) *Evolution* **39**, 783–791.
20. Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
21. Reynolds, J., Weir, B. S. & Cockerham, C. C. (1983) *Genetics* **105**, 767–779.
22. Cavalli-Sforza, L. L. & Piazza, A. (1975) *Theor. Popul. Biol.* **8**, 127–165.
23. Felsenstein, J. (1973) *Am. J. Hum. Genet.* **25**, 471–492.
24. Thompson, E. A. (1975) *Human Evolutionary Trees* (Cambridge Univ. Press, Cambridge).
25. Klein, R. G. (1989) in *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans*, eds. Mellars, P. & Stringer, C. (Edinburgh Univ. Press, Edinburgh), pp. 529–546.
26. Cavalli-Sforza, L. L. & Edwards, A. W. F. (1967) *Am. J. Hum. Genet.* **19**, 223–257.
27. Cavalli-Sforza, L. L. & Edwards, A. W. F. (1967) *Evolution* **21**, 550–570.
28. Kidd, K. K. & Sgaramella-Zonta, L. A. (1971) *Am. J. Hum. Genet.* **23**, 235–252.
29. Fitch, W. M. (1971) *Syst. Zool.* **20**, 406–416.
30. Kidd, K. K. (1973) in *Atti del Colloquio Internazionale sul Tema: L'Origine dell'Uomo* (Accademia Nazionale dei Lincei, Rome), Quaderno No. **182**, 149–174.
31. Ammerman, A. J. & Cavalli-Sforza, L. L. (1984) *Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ. Press, Princeton, NJ).
32. Ruhlen, M. (1987) *A Guide to the World's Languages* (Stanford Univ. Press, Stanford, CA).
33. Cavalli-Sforza, L. L. (1966) *Proc. R. Soc. London* **164**, 362–379.
34. Lewontin, R. C. & Krakauer, J. (1974) *Genetics* **74**, 175–195.
35. Robertson, A. (1975) *Genetics* **81**, 775–785.
36. Wright, S. (1931) *Genetics* **16**, 97–159.
37. Cavalli-Sforza, L. L. & Bodmer, W. F. (1971) *The Genetics of Human Populations* (Freeman, San Francisco).
38. Bowcock, A. M., Hebert, J. M., Mountain, J. L., Kidd, J. R., Kidd, K. K. & Cavalli-Sforza, L. L. (1991) *Gene Geography*, in press.