Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 6002–6006, August 1988
Evolution

Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data

(origin of modern humans/phylogenetic trees/paleoanthropology)

LUIGI LUCA CAVALLI-SFORZA*, ALBERTO PIAZZA†, PAOLO MENOZZI‡, AND JOANNA MOUNTAIN*

*Department of Genetics, Stanford University, Stanford, CA 94305; †Institute of Medical Genetics, University of Torino, Torino 10126, Italy; and ‡Institute of Ecology, University of Parma, Parma 4311, Italy

Contributed by Luigi Luca Cavalli-Sforza, May 12, 1988

ABSTRACT **The genetic information for this work came from a very large collection of gene frequencies for "classical" (non-DNA) polymorphisms of the world aborigines. The data were grouped in 42 populations studied for 120 alleles. The reconstruction of human evolutionary history thus generated was checked with statistical techniques such as "bootstrapping." It changes some earlier conclusions and is in agreement with more recent ones, including published and unpublished DNA-marker results. The first split in the phylogenetic tree separates Africans from non-Africans, and the second separates two major clusters, one corresponding to Caucasoids, East Asians, Arctic populations, and American natives, and the other to Southeast Asians (mainland and insular), Pacific islanders, and New Guineans and Australians. Average genetic distances between the most important clusters are proportional to archaeological separation times. Linguistic families correspond to groups of populations with very few, easily understood overlaps, and their origin can be given a time frame. Linguistic superfamilies show remarkable correspondence with the two major clusters, indicating considerable parallelism between genetic and linguistic evolution. The latest step in language development may have been an important factor determining the rapid expansion that followed the appearance of modern humans and the demise of Neanderthals.**

The reconstruction of human phylogeny from contemporary genetic information was first attempted (1–4) by the use of gene frequencies of 20 alleles from five major blood-group systems known from 15 populations. The genetic information from all genes was cumulated by calculating a "genetic distance" between pairs of populations. Two independent methods developed for the purpose were used to reconstruct the phylogeny, with very similar results. One of them was based on independence of evolution in the branches resulting after every fission, and the other on maximum parsimony; neither, however, can define an origin (a "root") for the tree. When only information internal to the data set is used, it is necessary to assume constant evolutionary rates for setting a root. When this hypothesis was superimposed on constructed trees, the root separated African plus European populations on one side and the rest of the world on the other. The later addition of more genes (5), including *HLA* (6), caused little change in the shape of the phylogenetic tree.

Many protein and enzyme polymorphisms were detected in the 1960s and 1970s by electrophoretic methods but were initially tested on few samples. By using only three populations (Africans, Europeans, and East Asians), Nei (7) was able to consider many more genes. He concluded that blood groups and enzyme polymorphisms gave different results with respect to the location of the root, with blood groups still showing greater similarity between Africans and Europeans than between Europeans and East Asians, thus confirming earlier results on the position of the root. With enzymes and proteins, however, Europeans were closer to East Asians than to Africans. These markers carried more statistical weight than blood groups, so that the complete data located the root between Africa and Europe plus Asia. The conclusion remained unchanged on extension to other populations (8), but the analysis did not include the rich set of *HLA* data.

Since that time, there have appeared results generated by DNA restriction analysis of mitochondrial DNA (9, 10), β-globin (11), the Y chromosome (12), and 44 nuclear gene markers (13, 14). With some contradiction, they tend to confirm the African/non-African split, but they are affected by biological or statistical weaknesses that will be discussed in a separate paper in which we will also present new DNA-marker data. The classical marker data used here confirm this conclusion but are much more abundant, thus allowing us to study human evolution in greater detail and to test agreement with other sources of evolutionary information, both archaeological and linguistic.

Evolutionary Analysis of Classical Genetic Markers

Materials and Methods. The literature data were collected in the course of preparation of an atlas of human variation. Selection of the present material was guided by the desire to study a representative sample of the world aboriginal populations, balancing the need to have as many genes as possible with the need to minimize the number of gaps in the gene × population matrix. Two genetic distances corresponding to different evolutionary models were used for comparing populations in pairs: (*i*) the most investigated one, a family (15) that also includes distances used in earlier papers (refs. 1–4; see also ref. 16); and (*ii*) Nei's standard genetic distance (16), always with correction for sample size. The two distances were highly correlated ($r = 0.86$) and the relation between them was of almost perfect proportionality except at short distances. We found it safe to use gene × population matrices that had gaps, provided these were not too frequent, by calculating distances between pairs of populations only for genes known in both of the populations being compared. Tests of this statement included comparison of results of principal components and tree analysis based on both an incomplete matrix and a complete subset of the same matrix, as well as extensive experiments of simulation of the effect of random gaps on principal component maps, which are highly related to the highest splits in the tree (17). Gaps in the matrix used for the tree of Fig. 1 were 23.7%.

A recently introduced statistical technique, the bootstrap (18), a resampling method for obtaining standard errors that are difficult to estimate directly, proved very useful. According to this method, genes actually used are randomly sampled

---

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.

Abbreviation: kyr, thousand years ago.

with replacement, generating a new matrix of genes × populations with the same number of genes as the original, but in which some genes are missing and others are repeated. The procedure is repeated a sufficient number of times ("bootstraps"), each time producing a new matrix. From each matrix a given statistic is calculated; its standard error is the standard deviation of the values taken by the statistic in the bootstrap samples. Felsenstein (19) suggested using the bootstrap to test the reproducibility of the sequence of splits in the tree.

Tree of 42 World Populations. The tree shown in Fig. 1 was generated by average linkage analysis (20) of 42 populations representing the world aborigines: 7 African, 5 American (natives), 5 Oceanian, 6 European, and the rest Asian including insular Southeast Asia. All values are average gene frequencies for all data found in the literature satisfying the criteria of being aboriginal, with little or no admixture, pooling populations geographically. When geographic pooling gave rise to potential heterogeneity, an ethnolinguistic criterion of classification was added. Six of the 42 groups were formed on the basis of linguistic affinity: Nilosaharan and Bantu in Africa; Samoyed and other Uralic language speakers living near the Ural mountains; Northwest Americans speaking northern and southern Na-Dene languages; North Turkic, i.e., Northeast Asian populations whose language belongs to this subgroup of the Altaic phylum; Southeast Indian, speaking Dravidian languages. The tree in Fig. 1 is slightly simplified with respect to the original one of 42 populations in that all Europeans that clustered compactly together (Basque, Dane, English, Greek, and Italian) were pooled to form one population. All analyses, however, were done on the full tree.

The first split in the tree separates Africans from non-Africans and is reproducible, given that in 84 out of 100 bootstrap trees the first split separated from all other populations a cluster containing at least the four "core" African populations (Pygmies, West Africans, Bantu, and Nilosaharan). In most of the 84 trees the African cluster also contains the other two sub-Saharan African populations (San and Ethiopians). When these two are not with the core Africans they tend to join the Caucasoid group.

The next bifurcation separates two major "superclusters," the first of which, Northeurasian, splits into (*i*) Caucasoids and then (*ii*) Northeast Asians plus Amerindians. The Northeast Asian cluster separates further into a small cluster of Arctic populations, including Eskimos, and a cluster including both East Asians and North Asians. Caucasoids form a fairly tight group consisting of 12 populations, 5 of which were pooled as "Europeans" in Fig. 1. Lapps leave the cluster in 32% of the bootstraps, joining Asian Arctic popu-

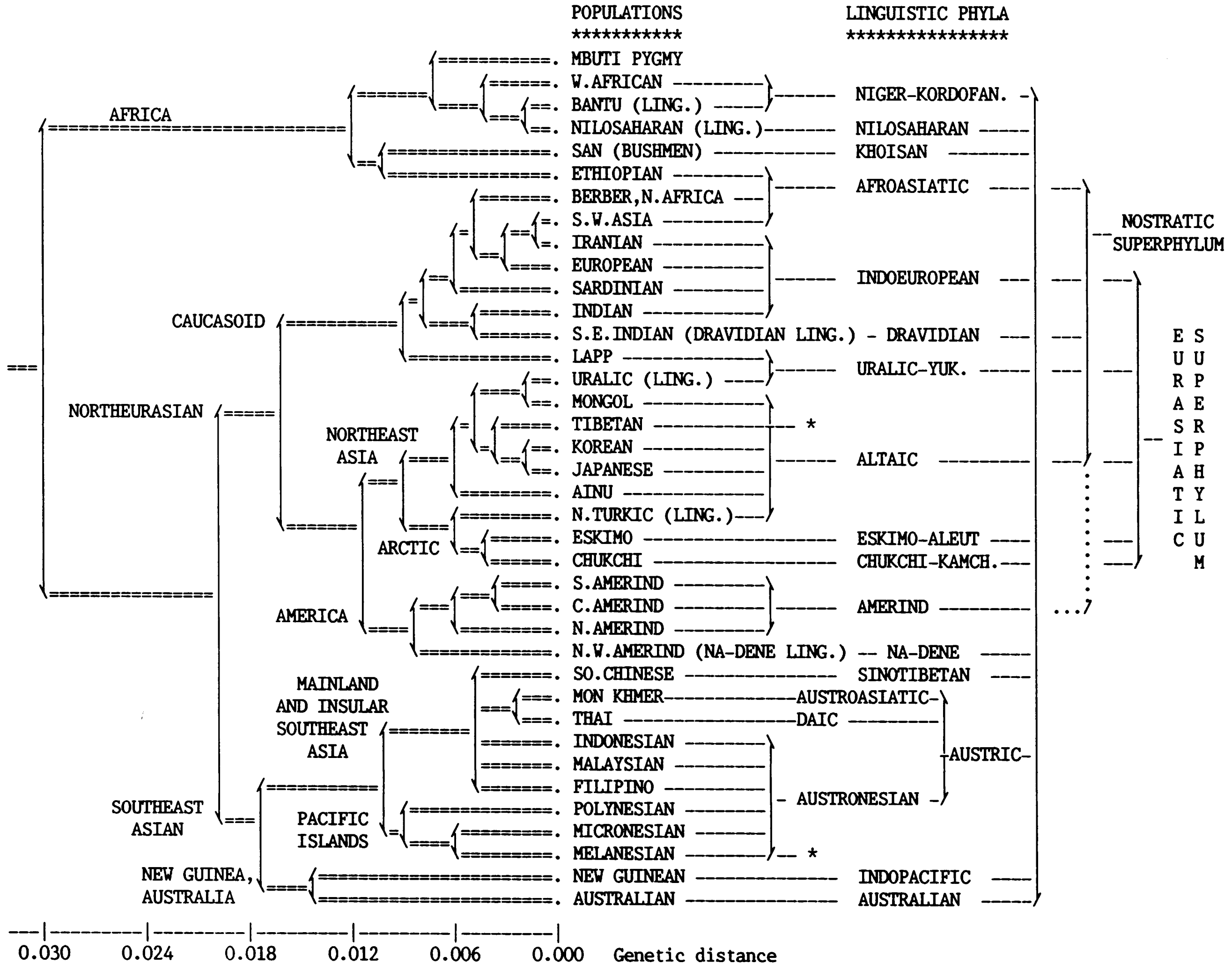

FIG. 1. Comparison of genetic tree and linguistic phyla. See text for details. (Ling.) indicates populations pooled on the basis of linguistic classification. The tree was constructed by average linkage analysis of Nei's genetic distances. Distances were calculated based on 120 allele frequencies from the following systems: *A1A2BO, MNS, RH, P, LU, K, FY, JK, DI, HP, TF, GC, LE, LP, PEPA, PEPB, PEPC, AG, HLAA* (12 alleles), *HLAB* (17 alleles), *PI, CP, ACP, PGD, PGM1, MDH, ADA, PTC, E1, SODA, GPT, PGK, C3, SE, ESD, GLO, KM, BF, LAD, E2, GM,* and *PG*.

lations. Berbers and Dravidians leave the Caucasoid cluster 20% of the time and tend to join respectively the African and one of the two major East Asian clusters. The Northeast Asian cluster is also reasonably compact, 81% of trees having at least 4 of the 6 populations shown in the tree in Fig. 1; most often lost are Tibetans (25%), Uralic speakers (13%), and Ainu (12%). The Arctic Northeast Asian cluster (Chuckchi, Eskimo, and North Turkic) is not very tight but is still well recognizable on bootstrapping. Amerinds are the tightest cluster, as in 79% of bootstraps all 4 populations are together, with Northwest Amerindians (speaking Na-Dene languages) being most easily lost (21%).

The Southeast Asian supercluster splits into (*i*) Southeast Asians proper (mainland and insular), a fairly tight cluster of the six populations seen in the tree of Fig. 1; on bootstrapping, Filipinos are lost 29% of the time, Malaysians 23%, and Indonesians 7%; (*ii*) the Pacific islanders, a cluster of three populations, not tight but clearly recognizable; and (*iii*) New Guineans and Australians, which remain together more than 50% of the time.

The earlier splits are all statistically significant by tests that will be described elsewhere. Of special interest is the second bifurcation shown in Fig. 1, separating Northeurasians from Southeast Asians. This split occurs most often among the bootstraps, but two alternative partitions are also fairly frequent: one separates Caucasoids from all Asian, Oceanian, and Amerindian populations, and the second separates New Guinean and Australian populations from all other non-African populations. We shall see later that the second bifurcation given in the tree receives support from an independent source as well.

Constant Evolutionary Rates. The evolutionary model (3, 4) on which phylogenetic tree analysis is based postulates that populations undergo fissions repeatedly over the course of time, dividing into two subpopulations that continue evolving independently after splitting and that may later split again. For the root to be established by tree analysis without the help of external evidence, the evolutionary rate in the various branches must be constant. The method of "average linkage" (20), used to construct Fig. 1, is based on the validity of this hypothesis, which can be tested more explicitly by the method of maximum likelihood. Unfortunately, this second method is virtually impossible with trees of 42 populations, but average linkage, which is computationally very rapid, gives results close to those of maximum likelihood (ref. 21 and citations in ref. 16). We used two approaches for checking the constancy of rates: a test of internal consistency, by evaluation of the treeness (17) with a procedure to be given in detail elsewhere, and a comparison with archaeological knowledge (next section). There are two major possible causes of deviation from constant evolutionary rate that may be avoided by appropriate selection and pooling of the samples.

(*i*) Populations that are or were of small size for long periods or that went through one or more severe demographic bottlenecks are affected by strong genetic drift and may show artificially long branches. We have therefore avoided as much as possible the use of individual populations, unless they were unique, and have pooled them with others of similar origin and used averages. A few unique, isolated populations, such as Mbuti Pygmies, San, and Lapps, were in any case the averages of many samples.

(*ii*) Mixed populations have shorter branches. Shorter branches cannot be appreciated in an average linkage tree, in which all branches are forced to be of equal size, but they could be shown by methods such as two cited in the Introduction (1–4). By a simple extension of the algebra used for the model of admixture shown in ref. 17, one can prove (as will be described elsewhere) that admixture between two branches of a tree, producing a third population, which then evolves independently from the two parental ones, has the following consequences: (*a*) the mixed population has a shorter branch; (*b*) the mixture attaches in the tree to the branch that has contributed the greater fraction; (*c*) the attachment takes place in a position that, in an average linkage tree, corresponds to an earlier time than that at which the admixture has taken place.

Bootstrapping, treeness tests, and independent methods of admixture analysis (22) agreed in giving evidence (to be presented elsewhere) that, among other populations, San and Ethiopians are, respectively, old and young admixtures of a majority of Africans and a minority of Caucasoids; similarly, Lapps are admixtures of a majority of Caucasoids and a minority of North Asians.

Correlations with Archaeology

There has been in the last few years considerable interest in the dating of the earliest anatomically modern humans, *Homo sapiens sapiens*. There seems to have developed some consensus about the validity of dates of early modern humans from the Border Caves and the Klasies River Mouth, both in South Africa, now dated at more than 100 kyr (100,000 years ago), with a date of 125 kyr for Laetoli in Tanzania (23) and 130 kyr for Omo I (24). In the rest of the world, findings are very poorly dated; however, a recent analysis of modern human remains from the Qafzeh cave in Israel suggests the date of 92 ± 5 kyr (25). This is about twice as old as the very approximate previous results. If the current archaeological data are accepted at face value, the origin of modern humans was in Africa and the expansion to the rest of the world started there. Naturally, considering the paucity of samples and the uncertainties in their dating, one cannot exclude the possibility that new archaeological discoveries may alter this picture.

The timing of the steps in the expansion can be very useful for calibrating the process of genetic differentiation and testing the constancy of evolutionary rates in our tree. Dates used for this aim in Table 1 are as follows.

(*i*) A time ≥92 kyr for the split between African and non-African has been matched with the genetic differentiation due to the first split of the tree.

(*ii*) The first entry to Australia took place at least 40 kyr (26), and the first settlement of New Guinea took place from Australia. We have matched this time with the node connecting Australia plus New Guinea with Southeast Asia (the third split in the tree).

(*iii*) The disappearance of Neanderthals and the first appearance of modern humans in southwestern Europe occurred 30–35 kyr, and somewhat earlier in Eastern Europe (27), for which data are less satisfactory. A time of 35 kyr was matched with the separation of Caucasoids from Northeast Asia.

Table 1. Comparison of genetic distances and archaeological time data

Clusters defining fission	Genetic distance (*G*)*	Time (*T*), kyr	*G*/*T*
African/non-African	29.7 ± 6.8	92	0.32 ± 0.07
Australian/S.E. Asian	18.4 ± 3.4	≥40	0.46 ± 0.09
Caucasoid/(N.E. Asian + Amerind)	16.6 ± 3.5	35	0.47 ± 0.10
N.E. Asian/Amerind	12.1 ± 1.8	15–35	0.81 ± 0.12 – 0.35 ± 0.05
N. Amerind/S. and C. Amerind	4.2 ± 1.0	15–35	0.28 ± 0.07 – 0.12 ± 0.03

*Nei's distances ×1000, with standard errors.

(*iv*) Two possible dates for the entry to America are 35 kyr and 15 kyr (28, 29). There seems to be more consensus for the second, later date, and in any case this is likely to have had greater demographic weight, given the relative number of sites. There is also uncertainty as to the best match in the genetic tree. Use of the fission between Northeast Asia and the Americas generates a genetic distance that is too large, if the tree has no direct descendants of the Northeast Asians who went to America. The fission between North Amerinds and Central plus South Amerinds generates a distance that is probably too small but that is at least uncomplicated by admixture with later arrivals to North America.

Table 1 shows the clusters defining the fissions listed above, the average genetic distances (G) between the clusters defining each fission, and the archaeological separation dates (T). If the evolutionary rates are constant, the G/T ratios should be constant. Leaving aside the Americas, for which there is uncertainty, we see that there is satisfactory agreement between the three values that are more dependable, as shown by standard errors. The older fission used for America seems to accord with the older date, and the younger one with the younger date. At least one can say that the American data are not inconsistent with the conclusions from the other three dates, from which the average G/T of 0.40 ± 0.05 was calculated. This is valid for long time intervals, for which there is the advantage that many distances are averaged and the distorting effects of admixture and drift are decreased and may even partially compensate for each other. For shorter intervals this G/T value should be used with caution.

Correlation with Linguistic Classification

There are approximately 5000 languages spoken today, and in a recent taxonomic effort they have been classified in 17 families or phyla (30). The phyla of the languages spoken by the populations studied in the genetic tree are listed in Fig. 1. Of the 17 phyla proposed by Ruhlen (30), only one, Caucasian, is missing for lack of adequate genetic data, but the limited genetic information available suggests that Caucasians are very similar to neighbors and would not generate anomalies if inserted into the tree. Inspection of Fig. 1 shows that every linguistic phylum corresponds to only one of six major genetic clusters defined by the tree. Exceptions are Mbuti Pygmies, who have lost their original language; Basques, who have kept their original language, which is an isolate (30); and Melanesians (starred in Fig. 1), who speak in part also Indopacific languages. More important exceptions are the following. (*i*) Ethiopians are classified genetically in the African cluster although they speak Afroasiatic languages, also spoken in North Africa and the Near East by people who are genetically Caucasoid. The evidence for genetic admixture of Ethiopians can explain the anomaly. (*ii*) Lapps associate linguistically with speakers of Uralic languages but genetically with Caucasoids and again have important genetic admixture. (*iii*) Tibetans (starred in Fig. 1) are associated genetically with the Northeast Asian cluster but linguistically with the Sinotibetan phylum, which is spoken in all of China. According to Chinese historians, the Tibetans originated from pastoral nomads of the steppes north of China; this origin explains their genetic association with the Northeast Asian cluster.

The correspondence between linguistic phyla and genetic clusters shows that they have similar origins, but phyla, being contained in the clusters, must have developed later. This suggests a time frame for their origin. Also of great interest are the "superfamilies" of languages recently proposed by some linguists. Greenberg (31) has classified all the many preexisting phyla of American native languages into three, which are incorporated in the Ruhlen classification shown in Fig. 1: a superphylum including all languages of Central and South America and many of North America, a phylum (Na-Dene) of languages spoken in the Northwest, and the Eskimo–Aleut phylum. It is interesting that the Austric superfamily postulated by Ruhlen (30) includes almost all of the Southeast Asian cluster, leaving out only the southern Chinese, who speak Sinotibetan languages.

Two other superfamilies have been suggested and are indicated at the extreme right of Fig. 1: Nostratic and Eurasiatic. The first follows a proposal by Soviet linguists (summarized in ref. 30) and includes six phyla that all belong to the Northeurasian major cluster; the sixth phylum, South Caucasian, is not given in Fig. 1 for reasons already discussed. The other superfamily, Eurasiatic, proposed by Greenberg and summarized by Ruhlen (30), overlaps but does not coincide with Nostratic; it includes other phyla also belonging to the Northeurasian cluster. A link of Nostratic with Amerind (dotted vertical line in Fig. 1) was recently suggested by Shevoroshkin (32). It is most striking that the union of Eurasiatic and Nostratic, with the Amerind extension, includes all, and only, the languages spoken in our major Northeurasian cluster, with the exception of Na-Dene, the origin of which is less clear. Greenberg (33) noted that the apparent contradictions between his Eurasiatic superfamily and the Russians' Nostratic superfamily can be resolved by considering time levels of separation.

Languages evolve more rapidly than genes. They can also undergo rapid replacement, even if the new language is imposed by an invading minority, provided this minority has adequate political and military organization, as in the "elite dominance" model (34). When this happens it may be difficult to find genetic traces of the invasion. Elites have developed only recently, however, rarely being older than 5000 years, and therefore episodes of rapid language replacement are relatively recent and often accounted for historically. In the more remote past, replacement was more rare, justifying the stability of the relation between linguistic phyla and genetic clusters.

The Process of Expansion of Modern Humans

Reconstruction of human evolution can be truly satisfactory only if information from all relevant sources of acceptable reliability gives a coherent answer. The present data and analysis offer an attempt at a detailed joint approach. The tree of Fig. 1 may be wrong in details; it would be surprising if a tree of this size based on current information were completely correct. Any change in the archaeological, genetic, or linguistic conclusions will require adjustment. Current agreement remains nevertheless very encouraging.

The model we use (1–4) clearly assumes expansion of modern humans from a nuclear area and replacement of other local preexisting populations, a model that has found resistance in some anthropological circles. The major evidence cited against the replacement model comes from the continuity of some traits in fossil crania from East Asia. This evidence has prompted Wolpoff *et al.* (35) to suggest an alternative model, which they call "the theory of multiregional evolution," based on an almost continuous population network (36) and standard clinal theory. But the multiregional model cannot provide an explanation of the most important phenomenon, the rapid expansion of modern humans to the whole Earth. Its inability to do so derives from its assumption that genetic populations are at equilibrium, whereas a rapid expansion is a disruption of a former equilibrium. Another hypothesis on which the model rests, the maintenance of potential interfertility among all living humans for very long periods, may be correct, but if interfertility can help to explain the local permanence of some traits, it does not help to understand the expansion of a new type.

First, one must explain the expansion. There are several examples of expansions, some easier to study because they took place in historic time (37). Archaeologists in the early 20th century postulated migrations in a facile way, often on the basis of inadequate evidence. Today they have reacted with a strongly antimigrationist stance, but they certainly do not deny all expansions; the important point is to find in every case whether there were conditions conducive to expansion and its maintenance (34). The Neolithic expansions were due to the introduction of new technologies of food production, allowing a substantial increase in the carrying capacity of the land. The introduction of farming in West Asia was the stimulus and the support for the Neolithic expansion to Europe and most probably in other directions (38). The introduction of farming in the Sahel, in addition to the introduction of iron technology, stimulated and supported the Bantu expansion to central and southern Africa (39). Pastoral nomadism, coupled with new social structures and with new techniques of transportation and warfare, mostly using the horse, supported the expansions in which the steppe nomads have been major actors until very recently and for several millennia (40). A rapid expansion can be viewed as a punctuationist (41) episode in evolution, and such events are likely to occur repeatedly in general and not only in human evolution.

Which stimuli determined, and which technologies helped, expansions of modern humans to the whole Earth? It seems very likely that an important role was played by a biological advantage that may have developed slowly over millions of years and undergone a final step only with the appearance of modern humans: a fully developed language. Isaac (42) has indicated archaeological evidence in favor of this hypothesis. From a speculative point of view, it seems reasonable that more efficient communication can improve foraging and hunting techniques, favor stronger social ties, and facilitate the spread of information useful for migratory movements. It also makes it easier to understand the rapid disappearance of Neanderthals, if they were biologically provided with speech of more modest quality than modern humans. In our society, until 100–150 years ago, deaf-mute people had very little chance of reproducing because of strong adverse social selection (43). Even if interfertility was potentially complete and there was little or no impingement, Neanderthals must have been at a substantial disadvantage at both the between- and the within-population level.

We thank J. H. Greenberg, F. C. Howell, K. K. Kidd, M. Ruhlen, W. S.-Y. Wang, and A. C. Wilson for reading our manuscript and providing useful comments. This work was supported by National Institutes of Health Grant GM 20467.

1. Cavalli-Sforza, L. L. & Edwards, A. W. F. (1963) in *Genetics Today, Proceedings of the 11th International Congress of Genetics, The Hague, The Netherlands*, ed. Geerts, S. J. (Pergamon, New York), No. 3, pp. 923–933.
2. Edwards, A. W. F. & Cavalli-Sforza, L. L. (1964) in *Phenetic and Phylogenetic Classification*, eds. Heywood, U. H. & McNeill, J. (Systematics Assoc., London), No. 6, pp. 67–76.
3. Cavalli-Sforza, L. L. & Edwards, A. W. F. (1967) *Am. J. Hum. Genet.* **19**, 233–257.
4. Cavalli-Sforza, L. L. & Edwards, A. W. F. (1967) *Evolution* **21**, 550–570.
5. Kidd, K. K. (1973) *L'Origine dell'Uomo*, Quaderno No. 182 (Accad. Naz. Lincei, Rome).
6. Piazza, A., Sgaramella-Zonta, L., Gluckman, P. & Cavalli-Sforza, L. L. (1975) *Tissue Antigens* **5**, 445–463.
7. Nei, M. (1978) *Jpn. J. Hum. Genet.* **23**, 341–369.
8. Nei, M. & Roychoudhury, A. K. (1982) *Evol. Biol.* **14**, 1–59.
9. Johnson, M. J., Wallace, D. C., Ferris, S. D., Rattazzi, M. C. & Cavalli-Sforza, L. L. (1983) *J. Mol. Evol.* **19**, 255–271.
10. Cann, R. L., Stoneking, M. & Wilson, A. C. (1987) *Nature (London)* **325**, 39.
11. 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.
12. Lucotte, G. & Ruffie, J. (1987) in *Second Congress International Historical Demography* (Inst. Natl. Etudes Demographique, Paris), in press.
13. Cavalli-Sforza, L. L., Kidd, J. R., Kidd, K. K., Bucci, C., Bowcock, A. M., Hewlett, B. S. & Friedlaender, J. S. (1986) *Molecular Biology of Homo Sapiens*, Cold Spring Harbor Symposia on Quantitative Biology (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
14. Bowcock, A. M., Bucci, C., Hebert, J. M., Kidd, J. R., Kidd, K. K., Friedlaender, J. A. & Cavalli-Sforza, L. L. (1987) *Gene Geogr.* **1**, 47–64.
15. Reynolds, J., Weir, B. S. & Cockerham, C. C. (1983) *Genetics* **105**, 767–779.
16. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
17. Cavalli-Sforza, L. L. & Piazza, A. (1975) *Theor. Popul. Biol.* **8**, 127–165.
18. Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans* (Soc. Industr. Appl. Math., Philadelphia).
19. Felsenstein, J. (1985) *Evolution* **39**, 783–791.
20. Sokal, R. R. & Michener, C. D. (1958) *Univ. Kans. Sci. Bull.* **28**, 1409–1438.
21. Astolfi, P., Kidd, K. K. & Cavalli-Sforza, L. L. (1981) *Syst. Zool.* **30**, 156–169.
22. Wijsman, E. M. (1984) *Hum. Genet.* **67**, 441–448.
23. Rightmire, G. P. (1984) in *The Origins of Modern Humans: A World Survey of the Fossil Evidence*, eds. Smith, F. H. & Spencer, F. (Liss, New York), pp. 295–325.
24. Day, M. H. & Springer, C. B. (1982) in *I Congres International de Paleontologie Humaine* (UNESCO, Paris), Vol. 2, 814–846.
25. Valladas, H., Reyss, J. L., Joron, J. L., Valladas, G., Bar-Yosef, O. & Vandermeersch, B. (1988) *Nature (London)* **331**, 614–616.
26. Jones, R. (1987) *Nature (London)* **328**, 666.
27. Howell, F. C. (1984) in *The Origins of Modern Humans: A World Survey of the Fossil Evidence*, eds. Smith, F. H. & Spencer, F. (Liss, New York), pp. xiii–xxii.
28. Bray, W. (1988) *Nature (London)* **332**, 107.
29. Fagan, B. M. (1987) *The Great Journey: The Peopling of the Ancient Americas* (Thames and Hudson, London).
30. Ruhlen, M. (1987) *A Guide to the World's Languages* (Stanford Univ. Press, Stanford, CA), Vol. 1.
31. Greenberg, J. H. (1987) *Language in the Americas* (Stanford Univ. Press, Stanford, CA).
32. Shevoroshkin, V. (1988) *Methods in Inter-Phyletic Comparisons* (Wiesbaden, F.R.G.), in press.
33. Greenberg, J. H. (1988) *Voprosy Jazykoznanija* (Moscow), in press.
34. Renfrew, C. (1987) *Archeology and Linguistics* (Cambridge Univ. Press, Cambridge, U.K.)
35. Wolpoff, M. H., Wu, X. & Thorne, A. G. (1984) in *The Origins of Modern Humans: A World Survey of the Fossil Evidence*, eds. Smith, F. H. & Spencer, F. (Liss, New York), pp. 411–483.
36. Weiss, K. M. & Maruyama, T. (1976) *Am. J. Phys. Anthropol.* **44**, 31–50.
37. Cavalli-Sforza, L. L. (1986) in *Evolutionary Perspectives and the New Genetics*, eds. Gershowitz, H., Rucknagel, D. L. & Tashian, R. E. (Liss, New York), pp. 13–30.
38. Ammerman, A. J. & Cavalli-Sforza, L. L. (1984) *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ. Press, Princeton, NJ).
39. Hiernaux, J. (1974) *The People of Africa* (Scribner, New York).
40. Zvelebil, M. (1980) in *The Cambridge Encyclopedia of Archeology*, ed. Sherratt, A. (Cambridge Univ. Press, Cambridge, U.K.), pp. 252–256.
41. Eldredge, N. & Gould, S. J. (1972) in *Models in Paleobiology*, ed. Schopf, T. J. M. (Freeman, San Francisco), pp. 82–115.
42. Isaac, G. (1972) in *Calibration of Hominid Evolution*, eds. Bishop, W. W. & Miller, J. A. (Scott. Acad. Press and Univ. of Toronto Press, Toronto), pp. 381–430.
43. Fraser, R. G. (1976) *The Causes of Profound Deafness in Childhood* (Johns Hopkins Univ. Press, Baltimore).