Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 6720–6722, July 1995

Commentary

Genetic support for the out-of-Africa theory of human evolution

Masatoshi Nei

Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, 328 Mueller Laboratory, University Park, PA 16802-5303

Currently a great controversy is going on over the origin of modern humans (*Homo sapiens*). One hypothesis termed the multiregional theory of human evolution (1) maintains that our ancestral species *Homo erectus*, whose brain size was considerably smaller than that of *H. sapiens*, moved out of Africa and spread to various parts of the world >1 million years (MY) ago and that *H. sapiens* evolved gradually from *H. erectus* worldwide with the effects of gene flow and natural selection. Yet this hypothesis asserts that several regional characters such as the shovel-shaped incisors in East Asians and the prominent brow ridge in Australian aborigines have remained unchanged for >1 MY, from the time of their ancestral species *H. erectus*. The other hypothesis called the out-of-Africa theory (2, 3) proposes that *H. sapiens* originated in Africa 100,000–200,000 years ago and that all present human populations outside sub-Saharan Africa are primarily descendants of a population that moved out of Africa ≈100,000 years ago.

The out-of-Africa theory was initially based on a phylogenetic analysis of restriction fragment length polymorphism data of mitochondrial DNAs (mtDNAs) sampled from different parts of the world (2) and some paleontological data (3). However, the mtDNA study was criticized by a number of authors. (*i*) The topology of the mtDNA tree was statistically unreliable (4), and it was difficult to prove that the African populations separated first from the rest of the human populations. (*ii*) The estimate of the time of the deepest split of mtDNA lineages had a large standard error, so the time estimate could be as old as 800,000 years, depending on the assumptions made. This made it difficult to distinguish between the two hypotheses of human origins. (*iii*) mtDNA is inherited as though it is a single gene. Therefore, it is extremely difficult to infer the phylogenetic tree of human populations from mtDNA variation. Vigilant *et al.* (5) examined this problem by using DNA sequence data for the control region (≈1000 bp) of mtDNA. However, their data were still insufficient to resolve the above problems (6–8). For this reason, a number of authors have emphasized the importance of studying many nuclear genes to resolve this controversy (9, 10).

In this issue, Goldstein *et al.* (11) adopt this latter approach by using microsatellite DNA polymorphisms and estimate the time of the deepest split of human populations. Microsatellite loci are segments of tandemly repeated DNA with a short repeat length, usually 2–5 nucleotides. Microsatellite loci show an extensive amount of polymorphism with respect to the number of repeats. Therefore, they are useful for studying the phylogenetic relationships of populations. Experimental studies have suggested that these polymorphisms are generated primarily by gain or loss of a repeat unit, though there are some exceptions. This means that the pattern of mutation can be described approximately by the stepwise mutation model (12) in population genetics. Goldstein *et al.* (13) and Slatkin (14) recently proposed a genetic distance measure between two populations [average square distance (ASD)], which increases linearly with evolutionary time when mutation-drift balance is maintained throughout evolutionary time. That is, the expected value of ASD is $V_X + V_Y + 2\beta t$, where V_X and V_Y are the variances of allele size in populations X and Y, respectively, and β and t are the mutation rate per locus and the number of generations since the two populations X and Y diverged.

The troublesome feature of this measure is the presence of the terms V_X and V_Y that make the variance of ASD large. To rectify this problem, Goldstein *et al.* (11) now propose a new distance measure designated as $(\delta\mu)^2 = (\mu_X - \mu_Y)^2$, where μ_X and μ_Y are the mean allele sizes in populations X and Y, respectively. The expectation of $(\delta\mu)^2$ is $2\beta t$, and thus $(\delta\mu)^2$ has a smaller variance than that of ASD and can be used for estimating the time of population divergence without the knowledge of V_X and V_Y. Therefore, $(\delta\mu)^2$ is a substantial improvement over ASD. If we know β from other sources, the divergence time between two populations can be estimated by $t = (\delta\mu)^2/(2\beta)$ generations or $T = gt$ years, where g is the generation time in years. However, it should be noted that $(\delta\mu)^2$ still has a large variance compared with some other distance measures, and thus a large number of microsatellite loci are necessary to obtain a reliable estimate of evolutionary time. Note also that this distance measure is usually inefficient in obtaining a reliable phylogenetic tree (topology). The virtue of $(\delta\mu)^2$ is its linear relationship with time so that it can be used for estimating evolutionary time.

Goldstein *et al.* (11) applied $(\delta\mu)^2$ to the data set of 30 microsatellite loci (10) to construct a phylogenetic tree for 14 worldwide human populations. The tree obtained was not very reliable, but the deepest root of the tree separated Africans from non-Africans, as in many other studies. Therefore, by using information available on the mutation rate for dinucleotide microsatellite loci (5.6×10^{-4} per locus per generation) and generation time (g = 27), they estimated the time of divergence between African and non-African populations. The estimate obtained is 156,000 years with a 95% confidence interval of 75,000–287,000 years.

This estimate is of the same order of magnitude as that obtained from gene frequency data for 62 protein loci (15). The latter estimate is 116,000 years with a confidence interval (2 standard errors in both directions) of 46,000–184,000 years. The agreement between the two estimates improves if we assume a shorter generation time than Goldstein *et al.* (11) did. For example, if we assume g = 20, which seems more reasonable than g = 27 for early human populations, T becomes 115,000 years with a confidence interval of 56,000–213,000 years. (The estimate from protein polymorphism data is not affected by generation time.) However, it should be noted that both of these estimates depend on a number of assumptions, such as no severe bottlenecks of population sizes and a reliable estimate of the mutation rate or the amino acid substitution rate, and it is difficult to evaluate the extent of uncertainty due to these factors at present. Therefore, the confidence intervals of the two estimates may be considerably greater than those given above.

Previously mtDNAs generated time estimates that had large standard errors, so that they were not useful to distinguish between the two hypotheses. However, the number of nucleotides examined for mtDNA has gradually increased, and Horai *et al.* (16) published the entire DNA sequences (≈16,500 bp) for three humans (one from Africa, one from Europe, and one from Asia) and four ape species (common and pygmy chimpanzees, gorilla, and orangutan). By assuming that the orangutan and the other apes diverged 13 MY ago, they then estimated the time of the

earliest split of the African and non-African mtDNAs as 143,000 ± 18,000 years ago. Since the splitting of mtDNA lineages is known to precede the population splitting (17), this estimate gives an upper bound for the population splitting. Even this estimate gives a 95% confidence upper bound of 179,000 years. At present, this seems to be the most reliable estimate of the upper bound of splitting time between African and non-African populations. At any rate, the three sets of genetic data give surprisingly similar estimates of splitting time.

What is the implication of these findings on the controversy over the multiregional theory vs. the out-of-Africa theory of human evolution? If we combine the estimates of splitting time from nuclear DNA and mtDNA, it seems that the separation of non-Africans from Africans occurred at most ≈200,000 years ago but probably ≈115,000 years ago. This is apparently incompatible with the multiregional theory, where the divergence should have occurred much earlier, though a small amount of gene flow, which is supposed to have occurred in later generations, should have reduced the genetic divergence to some extent. In contrast, these results are consistent with the out-of-Africa theory, because according to this theory the population that later formed Europeans and Asians apparently moved out of Africa ≈100,000 years ago. Currently, one of the oldest fossil remains of modern humans has been found in Israel (18), and this could represent an ancestral population that formed non-African populations (19, 20), though this view has not been universally accepted (21).

Of course, this does not mean that the out-of-Africa theory is correct. The advocates of the multiregional theory have never specified the extent of gene flow required between different geographical areas (mainly sub-Saharan Africa, Europe, Northeast Asia, and Southeast Asia) of the world. Therefore, one may argue that the extent of gene flow was sufficiently large so that the genetic depth of the first split of human populations looks shallow (21). However, if this is the case, it seems very difficult to maintain regional continuity of certain morphological characters for >1 MY. To maintain this regional continuity the extent of gene flow must be very low. If so, the age of the latest common ancestor of human mtDNAs should be much higher than the observed value (22). It can also be shown that if the extent of gene flow is extremely small, even adaptive mutations are unlikely to spread through the entire world population (23). Of course, one can argue that the characters that show regional continuity have had selective advantage in the region in which they occur and that the extent of gene flow has actually been substantial. This argument can certainly explain the regional continuity, but what is the biological basis of the selective advantage? What are the extents of selective advantage and gene flow that can explain both the regional continuity of morphological characters and the evolutionary change of *H. erectus* into *H. sapiens* at the same time? These are difficult questions to answer.

Furthermore, the multiregional theory is based on cranial and dental characters of human remains, and if the regional continuity of some of these characters is real (24, 25), how can we explain this continuity in the framework of the out-of-Africa theory? Did the shovel-shaped incisors in present East Asians and the prominent brow ridge in Australians evolve independently again in the same regions? If so, is it due to the adaptation to similar environments or did it occur by chance? Or was there some hybridization between *H. erectus* and *H. sapiens* in these regions? Until these problems are solved, the controversy over the two hypotheses will probably continue. Note that there are some anthropologists who question the regional continuity of morphological characters (26–28). Therefore, it is also important to reexamine the validity of regional continuity.

Another unsolved issue is concerned with the replacement of *H. erectus* by *H. sapiens* without appreciable gene admixture, as is suggested by mtDNA data. Multiregionalists contend that this is unlikely, because if different groups of people come into contact, some degree of hybridization almost always occurs (21). Supporters of the out-of-Africa theory argue that the replacement is possible and may have been similar to the recent spread of African killer bees in South and Central America (29). This debate is difficult to resolve because we do not have enough fossil records to estimate the population size of *H. erectus*. If the population size of Eurasian *H. erectus* was small around 100,000 years ago for some reason, the replacement without hybridization would have been possible, but otherwise some hybridization seems to have occurred. In this respect it is worth noting that species extinction occurs very often in evolutionary process (30). Neanderthals who inhabited Europe from 130,000 years ago to 35,000 years ago apparently became extinct. It is also believed that a few species belonging to the genus *Homo* or its immediate ancestral genus *Australopithecus* became extinct during the last 2 or 3 MY (31, 32).

At any rate, to resolve this controversy and other problems in human evolution, it will be important to clarify the evolutionary pathways of human populations. For this purpose we can use microsatellite DNA loci, which are abundant in the human genome. However, the genetic distance measure that is appropriate for estimating divergence time is not necessarily good for phylogeny construction. It is therefore recommended that the phylogeny be reconstructed by using such distances as the modified Cavalli-Sforza distance (D_A) (17). Since the size of a human population changes all the time, it is also recommended that a phylogenetic tree be constructed without the assumption of constant rate of evolution. If $(\delta\mu)^2$ is to be used for estimating divergence times, a large number of loci must be used since the variance is very large. Furthermore, for studying the evolutionary relationships of distantly related species such as humans, chimpanzees, and gorillas, DNA sequences are more useful than microsatellite loci.

It should also be emphasized that the mutational pattern of microsatellite DNA does not strictly follow the stepwise mutation model (33–36). According to one experiment, only 78% of mutations were due to gain or loss of a repeat, and other mutations involved changes of two or more repeats (33). Furthermore, the mutation rate for these loci has not been well established. Therefore, more detailed studies on the mutation pattern and the mutation rate seem to be necessary. This is particularly important because microsatellite loci will be used for the study of evolution in many different organisms in the future.

I thank Bob Eckhardt, Henry Harpending, Naoko Takezaki, and Ken Weiss for their comments on earlier versions of this article.

1. Wolpoff, M. H., Wu, X. Z. & Thorne, A. G. (1984) in *The Origins of Modern Humans*, eds. Smith, F. H. & Spencer, F. (Liss, New York), pp. 411–483.
2. Cann, R. L., Stoneking, M. & Wilson, A. C. (1987) *Nature (London)* **325,** 31–36.
3. Stringer, C. B. & Andrews, P. (1988) *Science* **239,** 1263–1268.
4. Maddison, D. R. (1991) *Syst. Zool.* **40,** 355–362.
5. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. (1991) *Science* **253,** 1503–1507.
6. Hedges, S. B., Kumar, S., Tamura, K. & Stoneking, M. (1992) *Science* **255,** 737–739.
7. Templeton, A. R. (1993) *Am. Anthropol.* **95,** 51–72.
8. Maddison, D. R., Ruvolo, M. & Swofford, D. L. (1992) *Syst. Biol.* **41,** 111–124.
9. Nei, M. & Roychoudhury, A. K. (1993) *Mol. Biol. Evol.* **10,** 927–943.
10. Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R. & Cavalli-Sforza, L. L. (1994) *Nature (London)* **368,** 455–457.
11. Goldstein, D. B., Ruiz-Linares, A., Cavalli-Sforza, L. L. & Feldman, M. W. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 6723–6727.
12. Ohta, T. & Kimura, K. (1973) *Genet. Res.* **22,** 201–204.

13. Goldstein, D. B., Ruiz-Linares, A., Cavalli-Sforza, L. L. & Feldman, M. W. (1995) *Genetics* **139,** 463–471.
14. Slatkin, M. (1995) *Genetics* **139,** 457–462.
15. Nei, M. & Roychoudhury, A. K. (1982) *Evol. Biol.* **14,** 1–59.
16. Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. & Takahata, N. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 532–536.
17. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
18. Valdes, H., Reyss, J. L., Joron, J. L., Vallades, G., Bar-Yosef, O. & Vandermeesh, B. (1988) *Nature (London)* **331,** 614–616.
19. Stringer, C. B. (1990) *Sci. Am.* **263,** 98–104.
20. Clark, J. D. (1992) *Philos. Trans. R. Soc. London B* **337,** 201–215.
21. Wolpoff, M. H. (1989) in *The Human Revolution*, eds. Mellars, P. & Stringer, C. B. (Princeton Univ. Press, Princeton), pp. 62–108.
22. Takahata, N. (1993) *Mol. Biol. Evol.* **10,** 2–22.
23. Takahata, N. (1991) *Genetics* **129,** 585–595.
24. Frayer, D. W., Wolpoff, M. H., Thorne, A. G., Smith, F. H. & Pope, G. G. (1993) *Am. Anthropol.* **95,** 14–50.
25. Wolpoff, M. (1994) *Evol. Anthropol.* **3,** 38–39.
26. Groves, C. P. (1989) in *The Human Revolution*, eds. Mellars, P. & Stringer, C. (Princeton Univ. Press, Princeton), pp. 274–285.
27. Lahr, M. M. & Foley, R. (1994) *Evol. Anthropol.* **3,** 48–60.
28. Lieberman, D. E. (1995) *Curr. Anthropol.* **36,** 159–197.
29. Cann, R. L. (1992) in *Continuity or Replacement: Controversies in Homo sapiens Evolution*, eds. Bräuer, G. & Smith, F. H. (Balkema, Rotterdam), pp. 65–73.
30. Raup, D. M. (1991) *Extinction: Bad Genes or Bad Luck?* (Norton, New York).
31. Simmons, E. L. (1989) *Science* **245,** 1343–1350.
32. Wood, B. (1992) *Nature (London)* **355,** 783–790.
33. Weber, W. & Wong, C. (1993) *Hum. Mol. Genet.* **2,** 1123–1128.
34. Shriver, M. D., Jin, L., Chakraborty, R. & Boerwinkle, E. (1993) *Genetics* **134,** 983–993.
35. Valdes, A. M., Slatkin, M. & Freimer, N. B. (1993) *Genetics* **133,** 737–749.
36. Di Rienzo, A., Peterson, A. C., Carza, J. C., Valdes, A. M., Slatkin, M. & Freimen, N. B. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3166–3170.