Genetic Data and the African Origin of Humans

S. A. Tishkoff *et al.* (*1*) provide an intriguing analysis of human genetic variation at the CD4 locus. We are concerned, however, that their data do not provide significant support for the estimate that modern humans first emerged from Africa in the last 100,000 years. It appears that a robust estimate of this migration time will require the use of numerous loci.

The estimate, which is predicated on a set of assumptions listed in their paper, depends in part on estimating the relative ages of the Alu deletion [Alu(−)] allele in African and non-African populations. Tishkoff *et al.* argue that among Africans, the frequency of Alu(−) chromosomes linked to the progenitor [90 base pair (bp)] specific short tandem repeat polymorphism (STRP) allele is given by $e^{-N_A\mu}$, where N_A is the age of the Alu(−) allele and μ is the STRP mutation rate. (They give an equivalent expression for non-Africans, in which N_B represents the time of migration out of Africa.) Under the assumption of no back mutations, this expression does give the *expected* frequency of the 90-bp allele on Alu(−) chromosomes. Because many of the individuals in the sample will have a shared ancestry, the alleles found in different individuals are highly correlated, and so an estimate based on this procedure may have an extremely high variance.

In estimating the age of the Alu(−) mutation, it is convenient to consider the problem in a coalescent framework (*2*). In this view, the individuals in a sample are related to one another by some ancestral tree (strictly speaking, this is ancestry at a specified locus). When a mutation occurs at some point on the tree, all the individuals who trace their ancestry through that point on the tree will carry that mutation (recall the assumption of no back mutation). This means that a mutation that occurs near the root of the tree will often be carried out by a large proportion of the sample.

We have investigated the problem of establishing a lower bound on $N_A\mu$, given the authors' observation that 34 out of 85 non-recombinant African Alu(−) chromosomes carry the progenitor allele [some of the Alu(−) chromosomes seemed to be descended from a single recombinant and were excluded from the original analysis]. Their estimate of the migration time out of Africa is crucially dependent on this lower bound. No detailed theory exists for finding such a bound analytically. We can, however, approximate confidence intervals with the use of simulations of the coalescent process. If the entire African sample consisted of non-recombinant Alu(−) chromosomes, it would be reasonable to set bounds on $N_A\mu$ by using standard coalescent assumptions to generate random trees with 85 tips (*3*).

In this case, we also know the Alu(−) frequency in the total sample, and it is possible to use the coalescent approach to get the distribution of mutations in the Alu(−) chromosomes *conditioned* on that frequency. In order to do this, we have generated trees of 806 chromosomes [the sample size in the article (*1*)] and selected only those that contain a clade of 132 chromosomes [the total number of Alu(−) chromosomes]. The relative times of the nodes in the tree of 806 are drawn from an exponential distribution [the parameter is $\binom{n}{2}^{-1}$ between nodes n and $n-1$] (*4*). Taking the clade of 132 to correspond to the 132 Alu(−) chromosomes, we have now specified the relationships within a simulated data set in which all the relative branch lengths have been drawn from the appropriate conditional distribution.

In the original data set, 47 of the 132 Alu(−) chromosomes were recombinants and were excluded from the analysis. In order to further condition our own analysis on this information, we have selected only those trees in which a clade of 47 lies within the clade of 132 Alu(−) chromosomes.

This procedure has allowed us to generate trees of 85 individuals whose relationships to the larger sample closely mimic those in the original data set. Each simulation specified the relative lengths of all the branches, and so picking a trial value of $N_A\mu$ for the top of the Alu(−) clade determined the expected number of mutations along each branch. For each simulated tree, and trial value of $N_A\mu$, the number of mutations on each branch was drawn from a Poisson distribution with that expected value.

Our results, based on 10,000 random trees that meet the above criteria, are rather striking. The value of $N_A\mu$, estimated using the method of Tishkoff *et al.*, is 0.916; however, we have found the lower bound to be 0.12 at the 5% level of statistical significance (*5*). This indicates that their estimate of $N_A\mu$ could be seriously in error, and even without taking into account the variance in the other estimates in the original calculation, the technique used cannot reject a migration time out of Africa as much as sevenfold greater than the original estimate.

It is also possible to analyze the estimate of $N_B\mu$ in a similar manner. In this case, however, the coalescent assumption of single origin might break down because there may have been numerous Alu(−) chromosomes in the proposed migration event.

A further problem with the argument presented by Tishkoff *et al.* (*1*) is that, in light of other studies of population coalescent times for mitochondrial DNA (*6*), the Y chromosome (*7*), and autosomal microsatellite markers (*8*), it seems unlikely that the Alu polymorphism is as old as 5 million years, as implicitly suggested. However, if it were this old, the absence of the Alu(−) allele in chimpanzees would not rule out an age of more than 5 million years. That is, the Alu(−) allele could have existed in the ancestral population of both humans and chimpanzees and subsequently been lost from the chimpanzee lineage.

This analysis shows that as a result of the shared ancestry of individuals in a population, estimates of mutation times—or population divergences based on a single mutating locus (STRP here)—can be highly unreliable, even when large samples of individuals are used.

Jonathan K. Pritchard
Marcus W. Feldman
Department of Biological Sciences,
Stanford University,
Stanford, CA 94305, USA
E-mail: jkp@charles.stanford.edu
marc@charles.stanford.edu

REFERENCES AND NOTES

1. S. A. Tishkoff *et al.*, *Science* **271**, 1380 (1996).
2. F. Tajima, *Genetics* **105**, 437 (1983).
3. The generation of random trees is described in detail by R. R. Hudson, in *Oxford Surveys in Evolutionary Biology*, D. Futuyma and J. Antonovics, Eds. (Oxford Univ. Press, Oxford, 1990), vol. 7, pp. 1–44.
4. This conditional sampling procedure borrows from an idea introduced for a different problem by R. R. Hudson, in *Mechanisms of Molecular Evolution*, N. Takahata and A. G. Clark, Eds. (Sinauer, Sunderland, MA, 1992), pp. 23–36. There is an implicit assumption of constant population size here; this will lead to a conservative lower bound on $N_A\mu$. However, it should be noted that the observation of both the Alu(−) mutation and the Alu(−) recombinant type being in high frequency argues against a model as extreme as exponential population growth [M. Slatkin and R. R. Hudson, *Genetics* **129**, 555 (1991)].
5. This result is based on at least 5% of the simulated sub-trees having a mutant frequency at least as high as that observed by Tishkoff *et al.* for trial values of $N_A\mu \geq 0.12$.
6. R. L. Cann, M. Stoneking, A. C. Wilson, *Nature* **325**, 31 (1987); L. Vigilant, M. Stoneking, H. Harpending, K. Hawkes, A. C. Wilson, *Science* **253**, 1503 (1991).
7. R. L. Dorit, H. Akashi, W. Gilbert, *Science* **268**, 1183 (1995); M. F. Hammer, *Nature* **378**, 376 (1995); P. N. Goodfellow, *ibid.*, p. 379.
8. D. B. Goldstein, A. R. Ruiz Linares, L. L. Cavalli-Sforza, M. W. Feldman, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6723 (1995).

26 April 1996; accepted 28 August 1996

Response: Distinguishing between the "Out-of-Africa" model and the "Multiregional"

model of human evolution depends on the demonstration of evolutionarily recent time depths for alleles found in non-African populations. Multiregional model enthusiasts also argue for an African origin, but place this origin at 1 million years before present (Y.B.P.), the approximate time at which *Homo erectus* remains can be identified outside Africa. Thus, for the "Out-of-Africa" model to be accepted, it is critical that allelic time depths be more recent than 1 million Y.B.P.

In support of the recent "Out-of-Africa" model, we (*1*) attempted to show that a single chromosomal segment, a CD4 locus haplotype composed of an Alu(−) allele and an STRP allele of 90 bp separated by 10 kb, had a recent time depth in non-Africans. As we emphasized in the article, in the absence of known recombination between the sites or mutation rates at the STRP marker, it is impossible to estimate an exact time of origin of this haplotype in non-Africans. However, by making certain conservative assumptions, it is possible to place likely upper bounds for this date. We used several methods of analysis to derive an upper bound for the coalescent date for non-Africans. One was based on the variance observed at the STRP on Alu(−) chromosomes outside versus inside Africa; this led to a date of 167,000 Y.B.P. Another analysis was based on the proportion of Alu(−) chromosomes with STRP alleles less than 110 bp outside versus inside Africa that carry the progenitor (90 bp) STRP allele. As an upper bound on this proportion, we examined its variability across five geographically diffuse sub-Saharan African populations that had more than 10 Alu(−) chromosomes. The proportion carrying the 90-bp repeat ranged from 0.25 in the Woloff to 0.53 in the Herero. We used 0.53 as an upper bound for this value across sub-Saharan Africa. For non-Africans, because of the small number of Alu(−) chromosomes not carrying the 90-bp allele, we assumed a Poisson distribution to obtain a lower 95% confidence bound for this number. With these two bounds, we obtained a maximum age of 313,000 Y.B.P. We also performed other conservative analyses [notes 40 and 41 in (*1*)], which gave additional estimates of maximal dates ranging to 450,000 Y.B.P.

All of these estimates of maximum age depend on the *assumption* that the Alu(−) allele has a maximum age of 5 million years and originated in Africa. This upper-bound estimate was used because the allele was not observed in chimpanzees or gorillas. Pritchard and Feldman state that the mutation could technically be even older, but they also agree that it is far more likely that this polymorphism is less than 5 million years old. A younger age seems likely because of the lifetime survival distribution for neutral mutations (*2*). In fact, our data argue for a more recent origin, albeit still ancient [note 42 in (*1*)]. Comparing variation in STRP allele size (calculated by any of several methods) shows that Alu(−) chromosomes have less variation than do Alu(+) chromosomes and are therefore likely to have a more recent coalescent.

Pritchard and Feldman use coalescent theory and a simulation to calculate a lower 95% confidence bound for $N_A\mu$. The sample of chromosomes on which their analysis is based derived from 10 extremely disparate African populations, spanning the entire continent, for which there must have been considerable relative endogamy. Such population structure would make more recent ages for the Alu(−) allele far less likely than would appear in Pritchard's and Feldman's simulation (*3*). Also, it is implausible that the population has been constant in size since the Alu deletion first occurred. Its rather high frequency in Africa suggests a rapid increase in the numbers of this chromosome soon after its introduction. Such growth would lead to a smaller estimate of variance for $N_A\mu$ than that calculated by Pritchard and Feldman.

Still, even under assumptions implausibly more conservative than ours, the upper bound for the estimate of the coalescent date of the Alu(−) chromosome in non-Africans is about 700,000 Y.B.P. (using Pritchard's and Feldman's estimate), still short of the 1 million years speculated by the "Multiregional" model. Their analysis thus supports our conclusion that a more recent date for an exodus of modern humans from Africa is more likely and that the CD4 data argue for the "Out-of-Africa" model rather than for the "Multiregional" model.

We originally stated (*1*) that the data we have obtained for the CD4 locus represent only a single realization of evolutionary history for Africans and non-Africans. As Pritchard and Feldman point out, it is tenuous to derive statistical distributions for coalescent times based simply on theory because of the arbitrary demographic assumptions required. The best way to derive such a distribution is empirically, combining the results of numerous different loci. Examination of linkage disequilibrium patterns for other systems in a fashion similar to what we have presented for CD4 should provide more definitive conclusions regarding the coalescence time for non-Africans.

Neil Risch
Department of Genetics,
Stanford University School of Medicine,
Stanford, CA 94305, USA
Kenneth K. Kidd
Sarah A. Tishkoff
Department of Genetics,
Yale University School of Medicine,
Post Office Box 208005,
New Haven, CT 06520, USA

REFERENCES

1. S. A. Tishkoff *et al.*, *Science* **271**, 1380 (1996).
2. N. Takahata and M. Nei, *Genetics* **124**, 967 (1990); N. Takahata, *Mol. Biol. Evol.* **10**, 2 (1993)
3. N. Takahata, *Genetics* **129**, 585 (1991).

3 October 1996; accepted 10 October 1996

Late Permian Extinctions

In their article "Comparative Earth history and Late Permian mass extinction" (*1*), A. H. Knoll *et al.* suggest that Late Permian extinctions were caused by the release to the atmosphere of massive quantities of carbon dioxide (CO_2) from the deep ocean; that the CO_2 buildup in the ocean resulted from primary production in the surface layer; and that, despite sluggish ocean circulation rates, the release of phosphorus from decaying organic matter in deep anoxic waters would have been sufficient to further stimulate photosynthesis (*2*), which would in turn have led to further organic decay (that is, positive feedback) before oceanic overturn and release of CO_2.

Knoll *et al.* otherwise deemphasize the role of nutrients in the Permian extinctions, but if ocean circulation had been sufficiently slow in the Late Permian, phytoplankton could have largely stripped the surface mixed layer of nutrients (*3*) so that a "nutrient collapse" could have occurred. Also, the expansion of gymnosperms during this time (*4*) and the greatly increased interior drainage associated with the formation of the Pangean supercontinent (*5*) could have sequestered large amounts of nutrients on land (*4*, *6*). Greatly decreased nutrient availability during the Late Permian is consistent with the loss of many suspension-feeding invertebrates and nekton and the differential survival of infaunal taxa that fed on organic-rich sediment (*6*, *7*, *8*), as described by Knoll *et al.* Moreover, before Late Permian extinctions, the Permo-Carboniferous was a time of increasing nutrient and food availability in the water column (*6*, *7*). Thus, just as global marine ecosystems were becoming increasingly dependent on greater food availability in the Late Paleozoic, the rug, so to speak, could have