Inferring the Geographic Origins of "Refugee" Chimpanzees in Uganda from Mitochondrial DNA Sequences

TONY L. GOLDBERG*

Harvard University, Department of Anthropology, Peabody Museum, 11 Divinity Avenue, Cambridge, MA 02138, U.S.A.

Introduction

Wild chimpanzee populations in Africa are rapidly declining due to deforestation, mining, and other practices that destroy chimpanzee habitat (Hamilton 1981; Struhsaker 1981; Hamilton 1984; Weber 1987; Teleki 1989; Howard 1991; Harcourt 1996). The hunting of chimpanzees for food and for trade on the international pet market exacerbates this trend. The pet trade has been particularly damaging; at least 10 chimpanzee infants die for every one that survives to its final destination (Teleki 1989). A secondary consequence of these threats is the proliferation of "refugee" chimpanzee populations. Increasing numbers of confiscated chimpanzees are being crowded into underfunded zoos and rehabilitation centers.

In most countries in which chimpanzees live, conservation resources are desperately scarce. Knowing the principal areas from which chimpanzees are being captured could help governments direct conservation efforts and distribute law-enforcement resources optimally. Determining the origins of captive chimpanzees could also help with reintroduction efforts. Reintroduction is generally problematic (Campbell 1980) and is relatively unexplored as a prospect for chimpanzees. Nevertheless, individuals could in theory be reintroduced to their natal populations if these could be determined. More realistically, entire captive populations could be relocated to islands or to ecologically suitable but unoccupied mainland habitats (e.g., forests from which chimpanzees have become extinct). Knowing the geographic origins of captive chimpanzees would be invaluable for assembling artificial populations from individuals of like provenance. Inbreeding and outbreeding depression could thus be reduced (Ralls & Ballou 1982; Templeton 1986), and the deleterious genetic effects of inadvertent immigration of reintroduced animals into natural populations could be minimized.

Unfortunately, the geographic origins of captive chimpanzees are notoriously difficult to determine (Teleki 1989). Morphology is unreliable even for classifying chimpanzees into subspecies (Reynolds & Luscombe 1971; Shea & Coolidge 1988; Uchida 1992). This study attempts to infer the geographic origins of 26 confiscated chimpanzees housed in Uganda's Entebbe Zoo through analysis of their mitochondrial DNA (mtDNA). Although reintroduction of individuals to their natal forests has not been considered for these animals, relocation of the entire population to an island habitat has. The Entebbe Zoo chimpanzees were confiscated between 1967 and 1994 from a variety of sources: 10 from traders in various Ugandan villages, 6 from traders in Uganda's capitol, 6 from Entebbe International Airport, and 4 from a traveling circus. Because of the tortuous routes by which many of the chimpanzees arrived in Entebbe, they could have originated from anywhere in equatorial Africa.

Methods

Hair samples were collected from the Entebbe Zoo chimpanzees during routine handling by zookeepers. Hair is a valuable source of genetic material for studies of wild chimpanzee populations and for chimpanzee conservation (Moore 1992; Morin et al. 1992; Morin et al. 1994). Hair was also sampled noninvasively (from sleeping nests) from 18 natural chimpanzee populations as part of a larger genetic survey of the easternmost chim-

*Current address: Department of Veterinary Pathobiology, College of Veterinary Medicine, University of Illinois, 2001 South Lincoln Road, Urbana, IL 61801, U.S.A., email tlgoldbe@uiuc.edu

Paper submitted August 5, 1996; revised manuscript accepted February 4, 1997.

Conservation Biology, Pages 1441-1446
Volume 11, No. 6, December 1997

panzee subspecies, *Pan troglodytes schweinfurthii* (Goldberg 1996; Goldberg & Ruvolo 1997). The natural populations sampled span a wide range of forest types within the geographic limits of *P. t. schweinfurthii* (Fig. 1.).

The DNA region chosen for study was a hypervariable segment of the mitochondrial control region. This region, corresponding to Anderson reference sequence coordinates 16042–16410 (Anderson et al. 1981), is the most quickly evolving region in primate mitochondrial genomes (Kocher & Wilson 1991). This, and the fact that mtDNA is maternally inherited, make the region ideal for detecting population differentiation in female-dispersing species like the chimpanzee. Also, the region has successfully been used to classify chimpanzees into subspecies (Morin et al. 1992). Laboratory methods are described elsewhere (Goldberg 1996). Nineteen published chimpanzee sequences from Tanzania were also used in the study (Morin et al. 1994).

Results and Discussion

A total of 281 DNA sequences were collected, yielding 123 unique haplotypes. These sequences are available through GenBank (accession numbers U77181–U77293) and are described in detail elsewhere (Goldberg 1996; Goldberg & Ruvolo 1997).

Figure 2 presents a phylogeny of haplotypes reconstructed from a matrix of distances generated by the computer program PHYLIP (Felsenstein 1993) with a maximum-likelihood distance correction and a transition:transversion ratio of 10:1. The algorithm used to generate the tree is the UPGMA algorithm (Sokal & Michener 1958). Alternate ratios of transition to transversion (as low as 1:1 and as high as 30:1) were also used, but they did not alter the topology. Neighbor-joining (Saitou & Nei 1987) and Fitch-Margoliash (Fitch & Margoliash 1967) tree-building algorithms did not yield significantly different trees. Maximum parsimony trees were not constructed because the number of variable positions (90) was less than the number of taxa (123), making parsimony analysis intrinsically weak (Swofford 1993; Rogers & Jorde 1995). The tree is rooted using three central African chimpanzees (*P. t. troglodytes*), three west African chimpanzees (*P. t. verus*), two bonobos (*P. paniscus*), and two humans (*Homo sapiens*) as outgroups. One western African chimpanzee sequence and both Bonobo sequences were taken from the literature (Horai et al. 1992; Morin et al. 1994), as were the two human sequences (Vigilant et al. 1991).

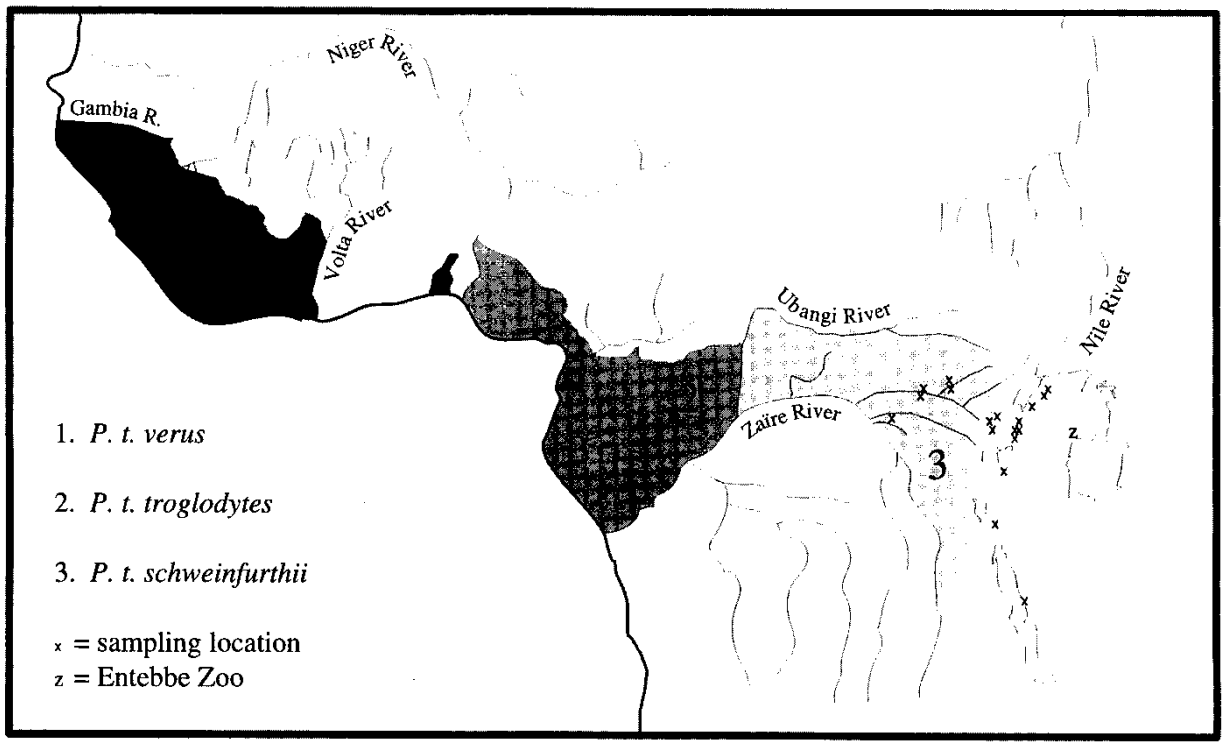

Figure 1. Approximate current distributions of chimpanzee subspecies in equatorial Africa. Locations of the 18 natural populations sampled and of Entebbe Zoo are marked.

*Figure 2. The UPGMA tree of 123 eastern chimpanzee haplotypes, rooted using human (Hsap), bonobo (Ppan), west African chimpanzee (Ptve), and central African chimpanzee (Pttr) as outgroups. Eastern chimpanzee haplotypes (*P. t. schweinfurthii*) are unlabeled, except for the 23 haplotypes identified in Entebbe Zoo, which are marked with asterisks. Branch lengths are not proportional to genetic distance.*

All of the Entebbe Zoo chimpanzees cluster with the easternmost subspecies. Comparison of the raw DNA sequences to published ones from known chimpanzee subspecies (Morin et al. 1994) confirm that the Entebbe Zoo chimpanzees are all *P. t. schweinfurthii.* These animals could not therefore have originated from central or western African locations within the ranges of the remaining two chimpanzee subspecies (Fig. 1). Haplotypes from natural populations (not labeled in Fig. 2) did not form exclusive clades on any tree generated.

Can the origins of these chimpanzees be further localized? Slatkin and Maddison (1989) describe a statistical method for detecting geographic clustering on phylogenetic trees that could help answer this question. Given a phylogeny of haplotypes, the geographic origins of which are known, one can reconstruct the minimum number of historical migration events needed to make the spatial distribution of haplotypes consistent with their inferred phylogenetic history. Geographic location is coded as an unordered, multistate character, and its evolution is traced along a tree reconstructed from nongeographic (genetic) information (Slatkin & Maddison 1989; Maddison & Maddison 1992). The minimum number of character state changes reconstructed along the tree represents the minimum number of inferred migration events, s. Nonparametric statistical probabilities can be assigned to s by comparing values from the reconstructed tree to values obtained from a large number of randomly generated trees of equal size. If observed values of s are lower than 95% of the randomly generated values, the null hypothesis may be rejected and the conclusion drawn that gene flow is restricted within the sample.

The Slatkin and Maddison technique is powerful in that it makes no assumptions about the underlying distribution of s values. The technique is, however, sensitive to the accuracy of the phylogeny used to generate the empirical s value. The UPGMA, neighbor-joining, and Fitch-Margoliash trees are therefore all included in the analyses that follow. Another limitation of the technique is that, in the event of a negative result, it cannot distinguish between migration and other forces that would also tend to obscure population differences. Homoplasy, for example, could artifactually increase the empirical s value even in a highly structured population, leading to a false acceptance of the null hypothesis. This is particularly worrisome in the present case because the mitochondrial control region evolves so quickly (Kocher & Wilson 1991). Given this bias, however, positive results (statistically significant s values) may be taken as strong evidence of restricted gene flow. Finally, as Slatkin and Maddison (1989) point out, determining actual migration rates from empirical s values requires many additional assumptions. The analyses that follow do not therefore attempt to estimate actual migration rates; rather, they use s in its raw form as a relatively reliable, if abstract, indicator of gene flow restriction.

The Slatkin and Maddison technique was applied to the UPGMA tree (Fig. 2) and to the neighbor-joining and Fitch-Margoliash trees (not shown). Three multistate, unordered geographic characters were defined. The first was an 18-state character representing each of the natural populations sampled in the study. The second was a 2-state character representing the biogeographic region from which the haplotypes came (eastern "insular" forests in Uganda, Rwanda, and Tanzania versus "mainland" forests in eastern Zaïre). The third was a 2-state character used to differentiate sequences found in Entebbe Zoo from sequences not present in this population.

For the location character, observed s values were lower than all values generated for the simulated null distribution, indicating that gene flow within the subspecies was significantly restricted at the population level (Fig. 3a; $p < 0.001$). Gene flow was similarly restricted at the regional level between Zaïrian and eastern forests (Fig. 3b; $p < 0.001$). Analysis of the Entebbe Zoo character (Fig. 3c) demonstrated that Entebbe Zoo haplotypes do not tend statistically to cluster on the UPGMA, neighbor-joining, or Fitch-Margoliash trees. Empirical values of s in each case were well within the 95% limits defined by the null distribution generated from 1000 random trees. In this respect, the Entebbe Zoo population is unlike the natural population sample or the regional sample.

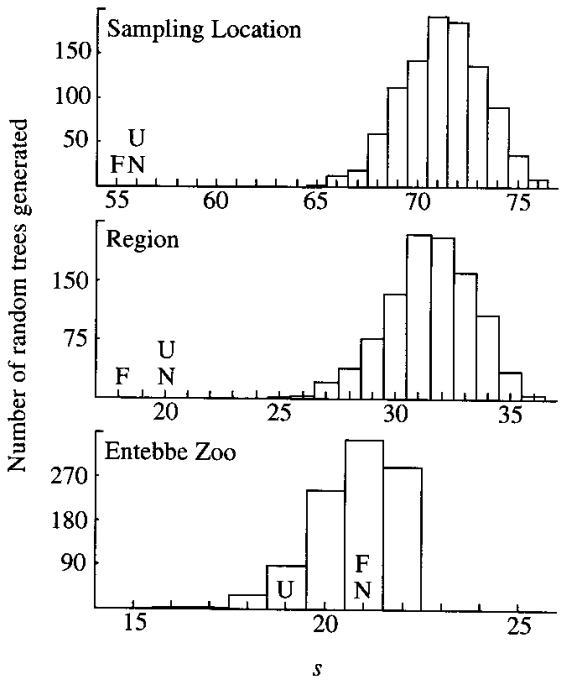

Figure 3. Null distributions of reconstructed migration events (s) based on randomly generated trees of 123 haplotypes. Histograms (open bars) represent the distribution of s values generated by tracing each character (sampling location, region, and Entebbe Zoo) along 1000 trees randomly generated by the computer program MacClade (Maddison & Maddison 1992). Empirical values of s inferred from UPGMA, neighbor-joining, and Fitch-Margoliash trees are indicated by letters (U, N, F, respectively) on each panel.

The natural population sample in the above analysis consisted of 100 haplotypes, whereas the Entebbe Zoo population consisted of only 23. The failure to detect clustering in the Entebbe Zoo sample could, therefore, have been a result of its smaller sample size. To test for this possibility, 1000 simulated populations were created in which 23 haplotypes from the natural population sample were randomly assigned to fictitious "zoo" populations. From the UPGMA, neighbor-joining, and Fitch-Margoliash trees, the reconstructed s value for each of these populations was determined as described above. The resulting distributions of 1000 s values gave sample-size-specific null distributions against which to compare observed s values for the real Entebbe Zoo population.

The null distribution obtained with the UPGMA algorithm had a range of $s = 10$ to $s = 32$, with a mean ($\pm$SE) of 20.8 $\pm$ 3.7. The distribution obtained by the neighbor-joining algorithm had a range of $s = 10$ to $s = 33$, with a mean of 20.7 $\pm$ 3.7. The distribution obtained using the Fitch-Margoliash algorithm had a range of $s = 10$ to $s = 33$, with a mean of 20.6 $\pm$ 3.5. In each case, the s value for the actual Entebbe Zoo population (19 and 21 respectively) was within 1 SE of the mean of its respective null distribution. In other words the s values for the Entebbe Zoo population were consistent with the hypothesis that Entebbe Zoo haplotypes were randomly sampled with respect to geographic origin.

In reality, the Slatkin and Maddison technique is sensitive enough to detect restrictions to gene flow in populations of sample sizes less than 23. To emphasize this point, the technique was applied separately to each of the 18 natural populations sampled. For each population, a two-state character was created to distinguish haplotypes found only in that population from haplotypes found outside of that population. An s value was then calculated for the UPGMA, neighbor-joining, and Fitch-Margoliash trees, and for 1000 randomly-generated trees. Observed s values were compared to simulated null distributions to detect gene flow restriction.

Sample sizes of haplotypes unique to each natural population were small (between 1 and 10). Nevertheless, 8 of the 18 natural populations individually showed significant s values. This result was consistent across all three tree-building algorithms. The 10 populations that did not show significant s values may have experienced extensive migration and/or homoplasy. Alternatively, gene flow restriction may not have been detected in these populations because of small sample sizes. In support of this hypothesis, levels of statistical significance obtained for populations in the previous analysis were negatively correlated with sample size ($r^2 = 0.44$, $p = 0.003$; 17 df). In other words, more extensively sampled populations tended to have more extreme s values. All populations with sample sizes greater than 7 showed significant s values. Given that the Entebbe Zoo sample size was 23, the probability that clustering was in fact present but not detected by the analysis is very low.

The most parsimonious interpretation of these results is that the Entebbe Zoo chimpanzees have not originated from any single localized area. Rather, Entebbe Zoo houses chimpanzees from widely dispersed locations within the range of *P. t. schweinfurthii*. In this respect, Entebbe Zoo is an unusual resource; few other captive populations likely consist exclusively of a single subspecies. In fact, the majority of captive chimpanzees are probably descendants of wild-caught *P. t. verus* (Teleki 1989). For this reason the Entebbe Zoo population is an attractive candidate for relocation to a wild or semi-wild habitat. The genetic consequences of inadvertent gene flow into or out of such a population would be minimal; major clades on the UPGMA, neighbor-joining, and Fitch-Margoliash trees consisted of haplotypes from widely dispersed natural locations anyway. But, nongenetic risks, such as the possibility of introducing exogenous infectious diseases, might be high (Dobson & May 1986; Atkinson 1989).

The results of this study could potentially be used to assign individual Entebbe Zoo chimpanzees to natal populations based on the naturally occurring haplotypes with which they tend to form local subclades on phylogenetic trees. The trees generated in this study would, by this method, suggest that 13 Entebbe Zoo chimpanzees originated from western Uganda, 2 from eastern Zaïre, and 1 from Rwanda. The provenances of the remaining 10 chimpanzees cannot be inferred. Such assignments can be made only with considerable uncertainty, however, given the many natural populations not sampled in this study. Increased sample sizes of both individuals and populations are clearly in order. Genetic analyses using more quickly evolving nuclear markers could also improve the accuracy with which natal populations are inferred.

More generally, the observation that the Entebbe Zoo chimpanzees do not come from any one location or region within the eastern subspecies range demonstrates that the "refugee" chimpanzee problem is one of cosmopolitan proportion. From the standpoint of law enforcement, finding one or several chimpanzee "supply centers" would have been favorable. Unfortunately, the results of this study indicate that halting the illegal capture of wild chimpanzees will be difficult, requiring the broad implementation of international laws that are also enforced at the local level.

Acknowledgments

I thank C. Manning and R. Wrangham for supervising hair collection from the Entebbe Zoo chimpanzees. M. Ruvolo provided laboratory facilities, and technical and scholarly advice. C. and L. Chapman, J. and T. Hart, T. P. Atwooki, K. M. Vital, and the faculty and staff of Makerere University Biological Field Station (Kanyawara, Uganda) and the Centre de Formation et de Recherche en Conservation Forestiere (Épulu, Zaïre) aided in the collection of samples from natural populations. Laboratory and field research was funded by the Wenner-Gren Foundation for Anthropological Research, the L. S. B. Leakey Foundation, the Charles A. and Anne Morrow Lindbergh Foundation, the Mellon Foundation, and the Department of Anthropology at Harvard University. Additional funding was received from Harvard University, the U.S. Department of Education, and the Cora Du Bois Charitable Trust. I am grateful to the Uganda National Council for Science and Technology and the government of Zaïre for permitting fieldwork.

Literature Cited

Anderson, S., et al. 1981. Sequence and organization of the human mitochondrial genome. Nature **290**:457–465.

Atkinson, I. 1989. Introduced animals and extinctions. Pages 54–69 in D. Western and M. C. Pearl, editors. Conservation for the twenty-first century. Oxford University Press, New York.

Campbell, S. 1980. Is reintroduction a realistic goal? Pages 263–269 in M. E. Soulé and B. A. Wilcox, editors. Conservation biology: an evolutionary-ecological perspective. Sinauer Associates, Inc., Sunderland, Massachusetts.

Dobson, A. P., and R. M. May. 1986. Disease and conservation. Pages 345–365 in M. E. Soulé, editor. Conservation biology: the science of scarcity and diversity. Sinauer Associates, Inc., Sunderland, Massachusetts.

Felsenstein, J. 1993. PHYLIP (phylogeny inference package), version 3.5c. Computer program distributed by the author. University of Washington, Seattle.

Fitch, W. M., and E. Margoliash. 1967. Construction of phylogenetic trees. Science **155**:279–284.

Goldberg, T. 1996. Genetics and biogeography of east African chimpanzees, *Pan troglodytes schweinfurthii*. Ph.D. thesis. Harvard University, Cambridge, Massachusetts.

Goldberg, T., and M. Ruvolo. 1997. Molecular phylogenetics and historical biogeography of east African chimpanzees. Biological Journal of the Linnean Society **61**:301–324.

Hamilton, A. C. 1981. The quaternary history of African forests: its relevance to conservation. African Journal of Ecology **19**:1–6.

Hamilton, A. C. 1984. Deforestation in Uganda. Oxford University Press, Nairobi.

Harcourt, A. H. 1996. Is the gorilla a threatened species? How should we judge? Biological Conservation **75**:165–176.

Horai, S., Y. Satta, K. Hayasaka, R. Kondo, T. Inoue, T. Ishida, S. Hayashi, and N. Takahata. 1992. Man's place in the Hominoidea revealed by mitochondrial DNA geneaology. Journal of Molecular Evolution **35**:32–43.

Howard, P. C. 1991. Nature conservation in Uganda's tropical forest reserves. World Conservation Union, Gland, Switzerland.

Kocher, T. D., and A. C. Wilson. 1991. Sequence evolution of mitochondrial DNA in humans and chimpanzees. Pages 391–413 in S. Osawa and T. Honjo, editors. Evolution of life: fossils, nucleotides and culture. Springer, Tokyo.

Maddison, W. P., and D. R. Maddison. 1992. MacClade: analysis of phylogeny and character evolution, version 3.0. Sinauer Associates, Inc., Sunderland, Massachusetts.

Moore, J. 1992. Chimpanzees, conservation, and hair. Pages 25–38 in Occasional papers 6. Museum of Anthropology, California State University Fullerton.

Morin, P. A., J. J. Moore, and D. S. Woodruff. 1992. Identification of chimpanzee subspecies with DNA from hair and allele-specific probes. Proceedings of the Royal Society of London **249**:293–297.

Morin, P. A., J. J. Moore, R. Chakraborty, L. Jin, J. Goodall, and D. S. Woodruff. 1994. Kin selection, social structure, gene flow and the evolution of chimpanzees. Science **265**:1193–1201.

Ralls, K., and J. Ballou. 1982. Effect of inbreeding on infant mortality in captive primates. International Journal of Primatology **3**:491–505.

Reynolds, V., and G. Luscombe. 1971. On the existence of currently described subspecies in the chimpanzee (*Pan troglodytes*). Folia Primatologica **14**:129–138.

Rogers, A. R., and L. B. Jorde. 1995. Genetic evidence on modern human origins. Human Biology **67**:1–36.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution **4**:406–425.

Shea, B. T., and H. J. Coolidge. 1988. Craniometric differentiation and systematics in the genus *Pan*. Journal of Human Evolution **17**:671–685.

Slatkin, M., and W. P. Maddison. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics **123**:603–613.

Sokal, R. R., and C. D. Michener. 1958. A statistical method for evaluating systematic relationships. University of Kansas Scientific Bulletin **28**:1409–1438.

Struhsaker, T. T. 1981. Forest and primate conservation in east Africa. African Journal of Ecology **19**:99–114.

Swofford, D. L. 1993. PAUP: phylogenetic analysis using parsimony, version 3.1. Illinois Natural History Survey, Champaign.

Teleki, G. 1989. Population status of wild chimpanzees (*Pan troglodytes*) and threats to survival. Pages 312–353 in P. G. Heltne and L. A. Marquardt, editors. Understanding chimpanzees. Harvard University Press, Cambridge, Massachusetts.

Templeton, A. R. 1986. Coadaptation and outbreeding depression. Pages 105–116 in M. E. Soulé, editor. Conservation biology: the science of scarcity and diversity. Sinauer Associates, Inc., Sunderland, Massachusetts.

Uchida, A. 1992. Intra-species variation among the great apes: implications for taxonomy of fossil Hominoids. Ph.D. thesis. Harvard University, Cambridge, Massachusetts.

Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes, and A. C. Wilson. 1991. African populations and the evolution of human mitochondrial DNA. Science **253**:1503–1507.

Weber, A. W. 1987. Socioecologic factors in the conservation of Afromontane forest reserves. Pages 205–229 in R. Mittermeier, editor. Primate conservation in the tropical rain forest. Alan R. Liss, New York.