Genetics and Biogeography of East African Chimpanzees (*Pan troglodytes schweinfurthii*)

A thesis presented

by

Tony Lawrence Goldberg

to

The Department of Anthropology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Anthropology

Harvard University
Cambridge, Massachusetts

May, 1996

© 1996 by Tony Lawrence Goldberg
All rights reserved.

Abstract

This study is a geographic survey of genetic variability in the easternmost subspecies of chimpanzee, *Pan troglodytes schweinfurthii*. The principal goal of the study is to test predictive hypotheses about the historical biogeography of the subspecies. Hypotheses are derived from paleoclimatological data and biogeographic theory. Forests in eastern Africa have expanded and contracted cyclically in response to oscillations in global climate throughout the Pleistocene. Biogeographic theory predicts 1) that chimpanzees were restricted to localized refugia during periods of minimal forest cover, and 2) that chimpanzees dispersed outward from such refugia when forests re-expanded during post-glacial climatic amelioration. Population genetic theory is used to predict the genetic consequences of such demographic events, both for the subspecies as a whole and for its constituent populations. The genetic hypotheses tested are 1) that eastern chimpanzees show evidence of a recent population-size bottleneck and 2) that the geographic apportionment of eastern chimpanzee genetic diversity is consistent with restriction to, and expansion from, Pleistocene forest refugia.

Chimpanzees were sampled from 19 geographically-defined populations in the countries of Uganda, Rwanda, Tanzania and Zaïre. Sampling locations were chosen to represent the full range of habitat types which eastern chimpanzees currently occupy. Sampling locations were also chosen to allow investigation of the potential roles of conspicuous topographic features (e.g. major rivers) as barriers to dispersal. DNA was collected non-invasively in the form of shed hair. Two hundred eighty-one mitochondrial control region DNA sequences were obtained. Eastern chimpanzees display levels of mitochondrial genetic variation which are low, and which are similar to levels observed in modern humans (*Homo sapiens*). "Mismatch distribution" analyses (Rogers and Harpending, 1992) reveal direct evidence of recent population expansion in the subspecies. These lines of evidence support the bottleneck hypothesis. However, insular

biogeographic analyses (MacArthur and Wilson, 1967) fail to support the hypothesis that chimpanzees were restricted to Pleistocene forest refugia prior to expansion. Population genetic diversity was unrelated to distance from refugia, forest area, or isolation. In addition, gene flow in the subspecies has been extensive, and relatively uninterrupted by phylogeographic barriers, such as major rivers or gaps in forest cover (Avise, 1987).

In conjunction, these lines of evidence suggest that the eastern chimpanzee population responded to Pleistocene climatic fluctuations by growing and shrinking in numbers only. The geographic range of the subspecies was unaffected by the cyclical contraction and expansion of forest into and out of refugia. This interpretation implies that chimpanzees have lived in dry, mosaic environments for much of their evolutionary history. These results are discussed in light of their implications for chimpanzee biogeography, evolution, and conservation. Appropriate comparisons with humans are made throughout the study. Additional analyses are presented which address the correlation between genetics and chimpanzee social behavior.

Acknowledgements

First and foremost, I thank my academic advisors. Richard Wrangham's unique blend of informed skepticism and carefully-administered encouragement left me confident in those ideas which survived the project's completion, and thankful for those which fell by the wayside. Maryellen Ruvolo's bag of laboratory and analytical tricks was bottomless, and kept me occupied during even the blackest phases of data collection and analysis.

Many thanks go also to my unofficial advisors. Colin and Lauren Chapman provided superhuman amounts of academic and personal support. David Pilbeam and Mark Leighton made invaluable comments on the dissertation itself as readers, and were helpful during countless less formal occasions. Other instrumental intellectuals were Irven DeVore, Alan Rogers, Vernon Reynolds, and Paul Ewald.

This project was generously 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. In addition, I thank the Uganda National Council for Science and Technology, and the government of Zaïre for supporting fieldwork. During the project, I received additional funding from Harvard University, the U. S. Department of Education, and the Cora Du Bois Charitable Trust.

For the fieldwork portion of this project, I owe my greatest thanks to Tuhairwe Peter Atwooki and to Katembo Mushgezni Vital, without whom I would still be lost in the forests of Uganda and Zaïre, respectively. Gilbert Isabirye-Basuta and the other faculty and personnel of the Kibale Chimpanzee Project in Uganda deserve special thanks for similar reasons. My undying gratitude goes to John and Terese Hart for ensuring my welfare while in Zaïre. Thanks also to "l'equipe du projet poile" in Zaïre for their enthusiastic help in all aspects of hair procurement. I am also grateful to Phil Harrell, Elizabeth Williamson, Christine Manning, Ellen Jacobs, Onyach Onecimo and Phil Morin

for collecting or otherwise providing hairs from places which I could not visit in person.

In the laboratory, Sarah Zehr, Miranda von Dornum and Anne Yoder provided invaluable technical advice, and a musically eclectic working environment. Additional helpful laboratory advice came from Phil Morin and Todd Disotell.

A number of people contributed indirectly to the completion of this project by providing intellectual, technical and emotional support. Specific thanks in this regard go to Rick Bribiescas for his generosity and unswerving distractibility (especially in matters piscatorial), and to Leah Gardner, Cheryl Knott and Jamie Jones. Others who contributed in similar ways include Todd Disotell, Joanna Lambert, Audrey Bribiescas, Libra Hilde, Tom Spang, Tom Hansen, Adrian Treves, Nathan Wolfe, Lynette Simon, Beryl Noble, Kathy Bolger, Meg Lynch, and Fran Boyle.

I could not have completed this project without the encouragement of my family, especially my parents Caren and Paul, whose unrelenting concern motivated me during all phases. A final note of thanks goes to my wife, Robin, for her love and support, and for so graciously agreeing to share me with East Africa's chimpanzees during the formative stages of our relationship.

Table of Contents

List of Tables and Figures

Tables

Figures

Chromosomatic and allied alterations are not the cause of evolution but only one of its manifestations. Proof of this is immediately furnished by the fact that chromosomal changes, modifications, etc., whether general or of detail, yield when plotted on a map patterns of distribution absolutely standard by comparison with those of entire taxonomic groups. In other words: the biogeography of the chromosome and all its bywords obeys the same laws as does that of the species, genus, family, etc. The former is accordingly but part of the whole to which also belongs the latter, and has no situation of privilege above or beyond it by space, time, form.

Léon Croizat, 1962
Space, Time, Form: The Biological Synthesis
p. 709

Chapter 1: Introduction

The field of anthropological genetics has benefited immensely from study of the chimpanzee, *Pan troglodytes*. *Pan* is genetically more similar to *Homo* than to any other primate genus (Sibley and Ahlquist, 1987; Caccone and Powell, 1989). Phylogenetically, chimpanzees are our closest living relatives. A wealth of independent molecular data sets now provide overwhelming statistical support for a hominoid phylogeny in which chimpanzees and humans cluster to the exclusion of gorillas (*Gorilla gorilla*), orang-utans (*Pongo pygmaeus*) and the lesser apes (Horai *et. al.*, 1992; Ruvolo, 1996a). This phylogeny is robust to high levels of intraspecific genetic variation (Ruvolo *et. al.*, 1994). It is probably the most hotly-debated yet strongly-supported hypothesis known to molecular systematics. It is therefore surprising that so little is known about the genetics of chimpanzees in their own right.

Chimpanzees are currently classified into three geographically-distinct subspecies (Hill, 1969; see Figure 1.1), following the recommendation of Schwartz (Schwartz, 1934). The westernmost subspecies, *P. t. verus*, historically ranged between the Gambia and Niger rivers in West Africa (Albrecht and Dunnett, 1971). Two geographic features demarcate the separation between *P. t. verus* and *P. t. troglodytes*, the central African subspecies: the Niger river, and the Dahomey gap. The latter is a historically-persistent and biogeographically-important gap in forest cover which divides the West African (Guinea) Forest Block and the Central African (Congo) Forest Block (Aubréville, 1949; Booth, 1954; Booth, 1957). *P. t. troglodytes* ranges between the Niger River and the Ubangi river, and as far south as the Zaïre River (Hill, 1969). The Ubangi River also defines the approximate northern and western range limit of the easternmost subspecies, *P. t. schweinfurthii* (Hill, 1969; Teleki, 1989). Its southern and southwestern range is demarcated by the Zaïre river, which also separates common chimpanzees from bonobos

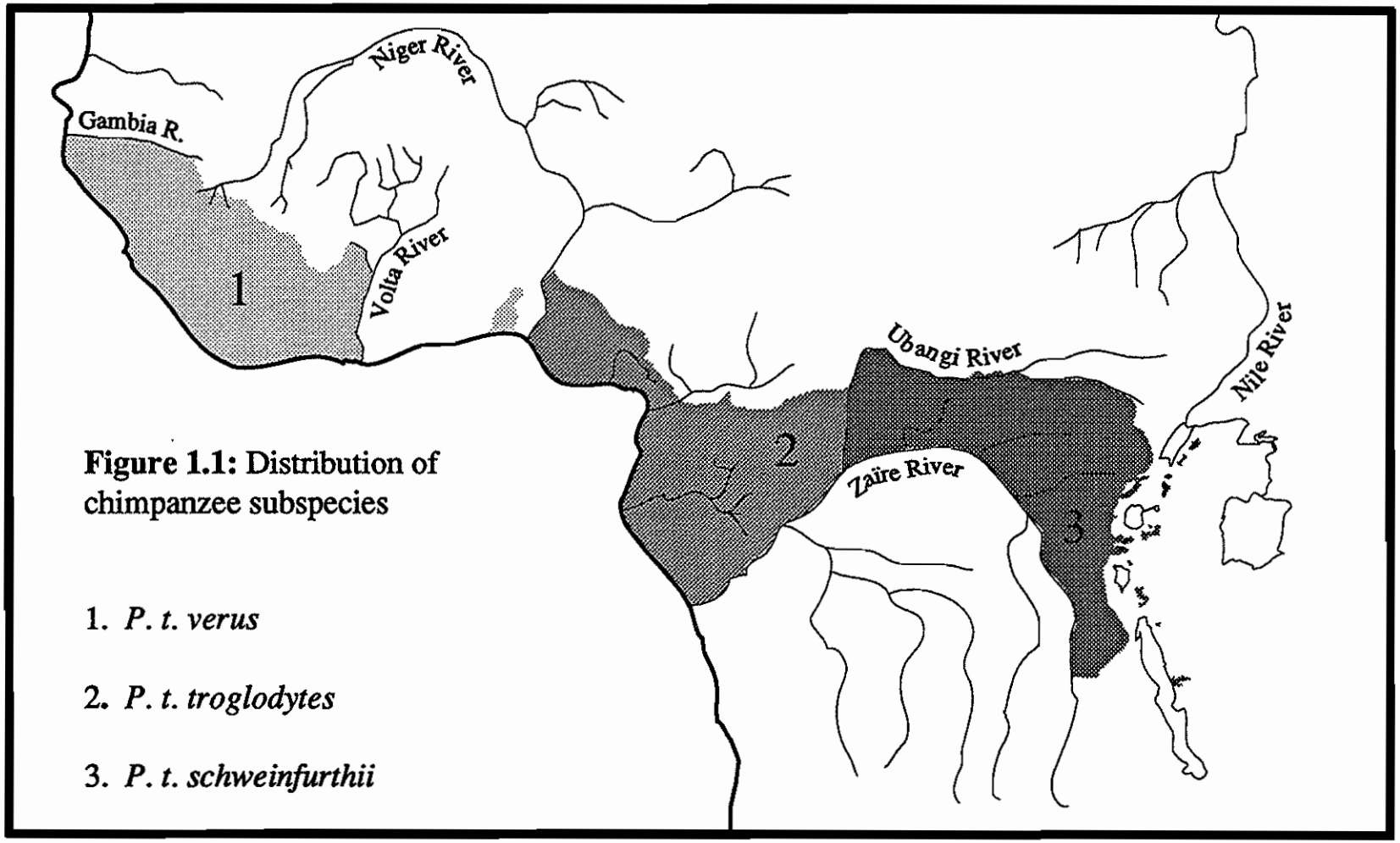


Figure 1.1: Distribution of chimpanzee subspecies

1. *P. t. verus*
2. *P. t. troglodytes*
3. *P. t. schweinfurthii*

(*Pan paniscus*; Hill, 1969; Teleki, 1989). *P. t. schweinfurthii* ranges as far east as the western Rift, which is coincident with the easternmost limits of suitable chimpanzee habitat (Goodall, 1986; Teleki, 1989). The Rift Valley and its associated lacustrine system may itself have been an important historical barrier to further eastward dispersal (Kortlandt, 1983).

Despite claims to the contrary (Reynolds and Luscombe, 1971), the three geographically-defined subspecies of chimpanzee are distinguished by morphological differences (Shea and Coolidge, 1988; Uchida, 1992). Although such morphological distinctiveness suggests genetic differentiation, it was not until recently that genetic divisions among the subspecies were documented (Morin *et. al.*, 1994a). Morin *et. al.* sequenced mitochondrial gene regions from the cytochrome b locus and the more rapidly-evolving control region locus in individuals from within each subspecies. Phylogenetic reconstruction demonstrated that the subspecies sort into distinct clades, the western subspecies being divergent with respect to the eastern and central subspecies. Calculations of times of divergence (calibrated to a 5 million year date for the divergence of *Pan* and *Homo*) imply that the westernmost subspecies has been isolated from the central and eastern subspecies for approximately 1.5 million years.

The current distribution of chimpanzee subspecies in Africa suggests that large rivers and gaps between blocks of forest have been critical in facilitating taxonomic differentiation within the genus *Pan*. It is therefore reasonable to hypothesize that similar geographic features have operated within chimpanzee subspecies. Morin *et. al.*(1994a) give intrasubspecific variability estimates for common chimpanzees, which suggest a rank-ordering of genetic variability of (in descending order) *P. t. troglodytes*, *P. t. verus*, *P. t. schweinfurthii*. Unfortunately, the geographic provenances of many of Morin *et. al.*'s samples are unknown, precluding geographic analyses on a finer scale. Furthermore, the DNA sequences are largely incomplete, leading to problems of alignment and potential

directional biases in estimates of variability within and between subspecies (pers. obs.). Without accurate information about the provenances of animals, and without systematic geographic sampling, it is difficult to test specific hypotheses about the influence of geography on genetics. Without complete genetic data, it is equally difficult to paint an accurate picture of genetic variability within any taxon.

This study attempts to provide the first geographically systematic genetic survey of an ape subspecies, the eastern subspecies of chimpanzee, *Pan troglodytes schweinfurthii.* The sampling area chosen for study lies between 26° and 32° longitude, and between 3° and -5° latitude (see Figure 1.2). Within this area, eastern chimpanzees inhabit two biogeographically different habitats. Medium to high altitude forests (e.g. Gombe, Kibale, Budongo), generally dominated by *Cynometra* and *Parinari* tree species (Howard, 1991), demarcate the easternmost limit of the chimpanzee range in Africa. These forests are "insular" in that wide gaps of woodland and savannah presently separate them (Langdale-Brown, Osmaston and Wilson, 1964; Hamilton, 1981; Howard, 1991). In the western portion of the study area, chimpanzees inhabit expansive lowland evergreen rainforests, predominantly in Zaïre. These forests are the main repository of species diversity for eastern African forest taxa (Hart, Hart and Murphy, 1989; Whitmore, 1990) and support remarkably diverse primate communities (Hart, Hart and Thomas, 1986; Thomas, 1991). The total area of suitable chimpanzee habitat in this region is larger by approximately two orders of magnitude (470,000 km^2) than the combined areas of all eastern insular forests (2550 km^2; Teleki, 1989). Major rivers transect the subspecies range within the Zaïrian forest block. These rivers are barriers to primate dispersal, and are known to have facilitated subspecies-level evolution in Colobine and Cercopithecine primates (Colyn, 1987; Colyn, 1991; Colyn, Gautier-Hion and Verheyen, 1991). Studies of New World primate distributions have also documented river effects, suggesting that fluvial barriers may generally be facilitators of taxonomic differentiation in primates (Hershkovitz, 1969;

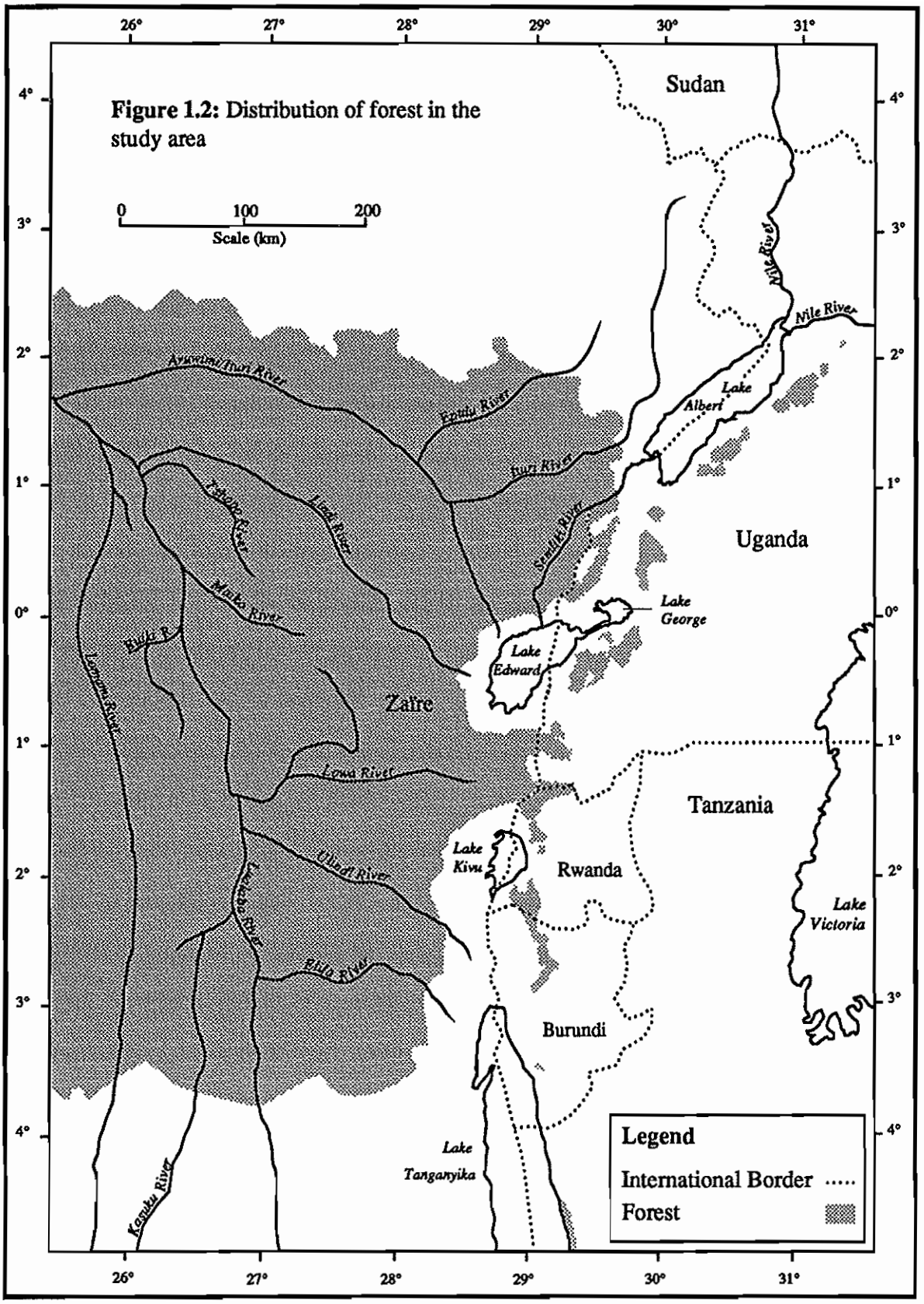


Figure 1.2: Distribution of forest in the study area
0
100
200
Scale (km)
Sudan
Uganda
Zaire
Rwanda
Burundi
Tanzania
Lake Albert
Lake George
Lake Edward
Lake Kivu
Lake Victoria
Lake Tanganyika
Nile River
Ituri River
Lowa River
Ulindi River
Rimo River
Kamiku River
Legend
International Border
Forest
26°
27°
28°
29°
30°
31°
4°
3°
2°
1°
0°

Ayres and Clutton-Brock, 1992).

The first goal of this study is to describe eastern chimpanzee genetic variability within this geographic context. The results provide an appropriate comparison for previous studies of humans (*Homo sapiens*), the best-studied hominoid to date. Studies of human genetic variability have been implicitly "geographic" in their traditional focus on racial/continental differences (Lewontin, 1972; Latter, 1980; Cann, Stoneking and Wilson, 1987; Stoneking *et. al.*, 1990; Merriwether *et. al.*, 1991; Vigilant *et. al.*, 1991; Maddison, Ruvolo and Swofford, 1992). Such studies have not, however, been able to place human genetic variability within a taxonomically-complete comparative context. Ape subspecies are likely to provide a good model for humans because of their comparable mobility and their similar life histories.

Rarely have human genetic data been collected or analyzed to test specific predictive hypotheses (Rogers and Harpending, 1992; Sherry *et. al.*, 1994; Rogers and Jorde, 1995). For example, the well-known "mitochondrial Eve" hypothesis of modern human origins (Cann, Stoneking and Wilson, 1987; Vigilant *et. al.*, 1991) was derived *a posteriori* from the empirical observation that levels of genetic diversity in humans were low, and from empirically-derived phylogenies which suggested an African origin for the species. Genetic tests of predictive hypotheses based on independent (e.g. archaeological) data have largely been retrospective (Wolpoff, 1989). The second goal of this study is therefore to form and test independently-derived predictive hypotheses about eastern chimpanzee genetic variability and its apportionment. These hypotheses are derived from biogeographic theory, and rely on the general notion that historical biogeographic processes have influenced the paleodemography of the subspecies. Specifically, biogeographic predictions are derived from Pleistocene refuge theory (Haffer, 1982; Mayr and O'Hara, 1986), Pleistocene paleoclimatological reconstruction (Emliani, 1955; Imbrie, 1985), and the theory of insular biogeography (MacArthur and Wilson, 1967).

In this sense, the study is less a survey of genetic diversity than a technologically-

enhanced inquiry into the historical biogeography of a single (particularly interesting) taxon. Because the data are genetic, the analytical techniques used are almost exclusively from the field of population genetics. However, because the study attempts to integrate a biological process (genetic evolution) with a geological explanation, it leans philosophically towards the (fortuitously-named) "panbiogeographic" approach of Croizat (Croizat, 1958; Croizat, 1964; Croizat, 1982; Brundin, 1988; Humphries *et. al.*, 1988).

The biogeographic model

The global climate during the last several hundred thousand years has been anything but stable. Dramatic shifts in global temperature have occurred throughout the Pleistocene (Emliani, 1955; Taylor *et. al.*, 1993), and have been globally synchronous (Williams, 1973; Imbrie *et. al.*, 1992). The current understanding of such shifts is based on the well-supported theory that large-scale climatic fluctuations result from oscillations in the Earth's orbital parameters (Milankovitch, 1930; Hays, Imbrie and Shackleton, 1976; Imbrie and Imbrie, 1980; Crowley *et. al.*, 1992; Imbrie *et. al.*, 1992; Liu, 1992). Vulcanism and other non-cyclic geographic processes have also influenced the Earth's climatic history (Kennett and Thunell, 1975; Bray, 1977; Raymo and Ruddiman, 1992). Global climatic reconstructions are most detailed for the past three hundred thousand years, where stratigraphic data have been accurately calibrated radiometrically (Hays, Imbrie and Shackleton, 1976).

Figure 1.3 is representative of the general pattern of global temperature change which has characterized the late Pleistocene. Isotopic data from other deep sea cores reveal strikingly similar patterns across a broad geographic range (e.g. Emliani, 1955; Emliani, 1966; Shackleton and Opdyke, 1973; Neftel *et. al.*, 1988; Van Campo *et. al.*, 1990; Eglinton *et. al.*, 1992; Imbrie *et. al.*, 1992; Thompson *et. al.*, 1995). The general pattern shown is also robust across a range of independent lines of isotopic evidence

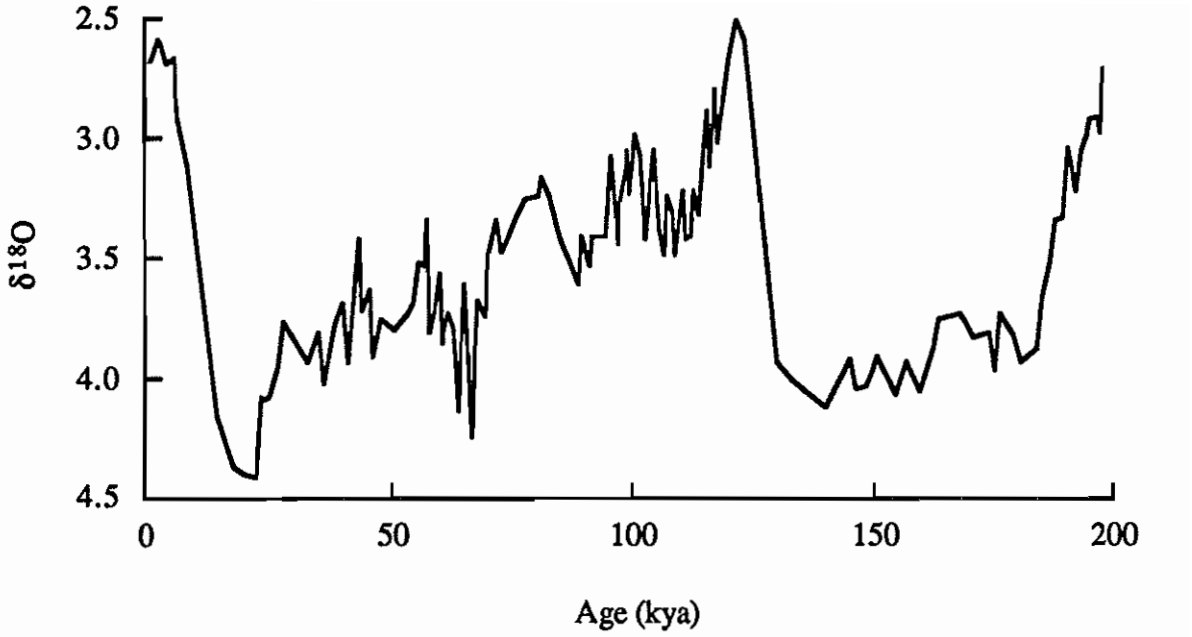


Figure 1.3: Oxygen isotope record for the past 200,000 years. The data were taken from Imbrie et. al. (1992), and were downloaded from the NOAA paleoclimatic data base (ftp://ftp.ngdc.noaa.gov/paleo/paleocean/specmap/specmap2/chn82i.stretch). The curve represents benthic $\delta^{18}O$ (in per mil) in a stratigraphically-calibrated mid-Atlantic deep sea core (CHN 82 24). Because $\delta^{18}O$ correlates negatively with global ice volume, the curve approximates global temperature and global humidity. The calibration method used results in an average estimated uncertainty of $\pm$ 2.5 ky for each climatic event. The general pattern shown in the figure is supported across a wide range of geographic locations and independent lines of isotopic evidence (Imbrie *et. al.*, 1992).

(COHMAP Members, 1988; Shackleton, West and Bowen, 1988; Imbrie *et. al.*, 1992). Periods of maximal global temperature occurred approximately 12,000 and 130,000 years ago. Prior to these periods, the global temperature was significantly colder. The transition from "ice age" conditions to warm conditions was rapid in each case, with changes of as much as 2.5° C occurring over periods as short as 300 years (Eglinton *et. al.*, 1992). This pattern is supported independently by studies of atmospheric dust (Taylor *et. al.*, 1993; Thompson *et. al.*, 1995), atmospheric carbon dioxide (Jasper and Hayes, 1990; Leuenberger, Siegenthaler and Langway, 1992), pollen (Guiot *et. al.*, 1989; Lezine and Casanova, 1991; Thompson *et. al.*, 1995), and lake and sea level changes (Washbourn, 1967; Street and Grove, 1979; Gallup, Edwards and Johnson, 1994).

Periods of low global temperature correspond to periods of enhanced glaciation, most significantly in the northern hemisphere (CLIMAP Project Members, 1976; Hays, Imbrie and Shackleton, 1976; Imbrie and Imbrie, 1980; Liu, 1992). Because large quantities of atmospheric water become trapped in ice sheets, globally cool periods tend also to be relatively dry (Williams, 1973; Imbrie, 1985). Periods of high global temperature cause diminution of the continental ice sheets and concomitant release of water into the oceans and atmosphere. Globally warm periods tend, therefore, to be relatively wet. The correlation between glacial phases and global climatic aridity is not, however, perfect or simple (Imbrie, 1985). Nevertheless, long periods of consistently low temperature correspond predictably with periods of global aridity. This observation has been confirmed directly by measures of ancient sea levels (CLIMAP Project Members, 1976; Gallup, Edwards and Johnson, 1994).

The local climate in tropical Africa has closely tracked these global changes (Livingstone, 1975; Moeyersons and Roche, 1982 ; Partridge, 1993; deMenocal, 1995). Isotopic data from from equatorial African lacustrine sediments parallel data from deep oceanic cores (Cerling, Hay and O'Neil, 1977; Gasse *et. al.*, 1990; deMenocal, 1995).

Periods of maximum East African alpine glaciation (e.g. on Kilimanjaro, the Rwenzori) coincide with periods of minimal global temperature (Livingstone, 1962; Osmaston, 1967; Downie and Wilkinson, 1972; Hamilton and Perrott, 1978). East Africa was predictably arid during globally cold periods (Bonnefille, Roeland and Guiot, 1990). Lake levels in East Africa were considerably depressed (Washbourn, 1967; Butzer *et. al.*, 1972; Street and Grove, 1979). Aeolian landforms (dunes) occurred in areas presently too moist to have allowed their formation (Grove and Warren, 1968; Burke, Durotoye and Whiteman, 1971; Sarnthein, 1978).

Local climatic changes have caused corresponding expansions and contractions of East African forests (Kendall, 1969; van Zinderen Bakker and Coetzee, 1972; Moeyersons and Roche, 1982). The timing and nature of these events have been documented in detail by a wealth of palynological studies from high-altitude lakes and mires in Kenya (Coetzee, 1964; Vincens, 1986; Coetzee, 1987; Hamilton, 1987), Uganda (Morrison, 1961; Hamilton, 1982; Hamilton, 1987; Taylor, 1990; Taylor, 1993), Rwanda (Deuse, 1966; Hamilton, 1982), and Burundi (Bonnefille and Riollet, 1988; Bonnefille, Roeland and Guiot, 1990; Jolly and Bonnefille, 1991). These studies have all documented the same basic trend, with only slight inter-regional variation in the timing of events. Montane environments between approximately 70 and 13 kya (Würm glacial period) were generally characterized by dry scrub vegetation. This vegetation was rapidly replaced by montane forest approximately 12.5 kya. Forests reached their maximum extent between approximately 10 and 8 kya. Between 2.5 kya and the present, forest retreat has accelerated due to human-induced deforestation (Phillipson, 1977; Hamilton, 1984; Hamilton, Taylor and Vogel, 1986; Howard, 1991). Low-altitude East African lake cores show the same trends for lowland forests (Kendall, 1969; Vincens, 1989; Sowunmi, 1991; Vincens, 1991). Archaeological evidence from a Zaïrian cave site (Matupi) suggests that parts of eastern Zaïre, today covered by evergreen lowland forest, were savannah at 22 kya (Van Neer, 1984). Preliminary evidence suggests that forest contractions and expansions

were coincident across tropical Africa (Maley, 1987; Elenga, Vincens and Schwartz, 1991; Maley *et. al.*, 1991).

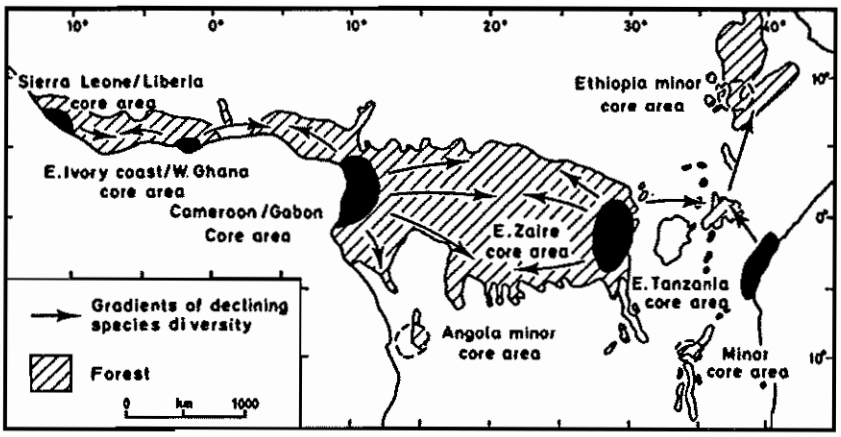
During periods of minimum global temperature and minimum overall forest cover, forests in Africa were restricted to areas where local conditions allowed their persistence despite general climatic aridity. These areas are the well-known biogeographic "refugia" (Hamilton, 1976; Endler, 1982; Haffer, 1982; Mayr and O'Hara, 1986; Hamilton, 1988). Refugia have been invoked to explain the observed disjunct distributions of many African forest taxa, from invertebrates to mammals (Lonnberg, 1929; Carcasson, 1964; Moreau, 1965; Moreau, 1966; Verdcourt, 1972; Hamilton, 1976; Diamond, 1979; Kingdon, 1981; Grubb, 1982; Chapman, 1983; Colyn, 1991). Indeed, the locations of refugia have been inferred retrospectively from data on the distributions of these taxa. Specifically, refugia have been reconstructed as areas which presently contain a high diversity of species (Brown, 1988) and large numbers of endemic taxa (Major, 1988). High diversity and endemism are expected in refuge areas under the assumptions that forest fragmentation facilitates allopatric speciation, and that dispersal of forest taxa from refuge areas is limited (Haffer, 1982; Mayr and O'Hara, 1986).

In eastern Africa, two forest refuge types have been described: montane refugia and lowland refugia. Geologically, the persistence of montane forest in refuge areas is explained by its retreat to lower elevations during cold, arid periods (Coetzee, 1964; Osmaston, 1967; Taylor, 1993). Initial reconstructions of lowland refugia were unable to provide a similar geological explanation for the localized persistence of lowland forest. On the basis of generally high species diversity and endemism, the principal eastern African lowland forest refuge was placed loosely in the topographically-undistinguished area of northeastern Zaïre (Hamilton, 1976; Grubb, 1982). Reanalyses of primate and small mammal distributions by Colyn (Colyn, 1987; Colyn, 1991; Colyn, Gautier-Hion and Verheyen, 1991) led to a redefinition of the northeastern Zaïrian refugium. According to

Colyn, a series of interconnected "fluvial" refugia were localized around the large rivers which transect Zaïre's forests. The area of forest represented by these refugia was as little as 6% of the area of present-day forests in the region (Colyn, 1991; Ruvolo, 1996b). The generally high species diversity observed in the northeastern Zaïrian region represents, under the Colyn model, a zone of mixing between endemics from the Zaïrian fluvial refuge and endemics from the montane refugia of the western Rift (see Figure 1.4).

Non-refuge forests extend from the eastern border of the Zaïrian forest block to the East African coast (Kingdon, 1989). These forest "islands" span a range of sizes, degrees of isolation, and ecological compositions (Kingdon, 1981; Struhsaker, 1981; Weber, 1987; Howard, 1991). Although gaps separating these forests have been enhanced by recent human forest clearance (Hamilton, 1984; Taylor, 1990), they may have been important historically as well. Ugandan forests separated by even small distances can have markedly different species compositions (Howard, 1991). Certain "conspicuous absences" reaffirm this notion. The red colobus subspecies *Colobus badius tephrosceles* is present in Uganda's Kibale Forest, but not in Itwara Forest only 15 km away (Struhsaker, 1975). Mountain gorillas (*Gorilla gorilla berengii*) occur in the Bwindi Forest of southern Uganda, but not in the ecologically-similar high-altitude forests of the nearby Rwenzori mountains (Howard, 1991).

The insular biogeographic model proposed by MacArthur and Wilson (MacArthur and Wilson, 1967) provides a useful framework for analyzing the species compositions of these forests. Under the MacArthur-Wilson model, the species diversity of an island (or an insular habitat) results from a balance between rates of extinction and rates of colonization. The theory predicts that island area should correlate positively with species diversity. Large islands support many species because of their relatively high microenvironmental heterogeneity. Large islands are also more likely to be encountered by dispersing colonists. The distance of an island from the "mainland" should, according to the theory,



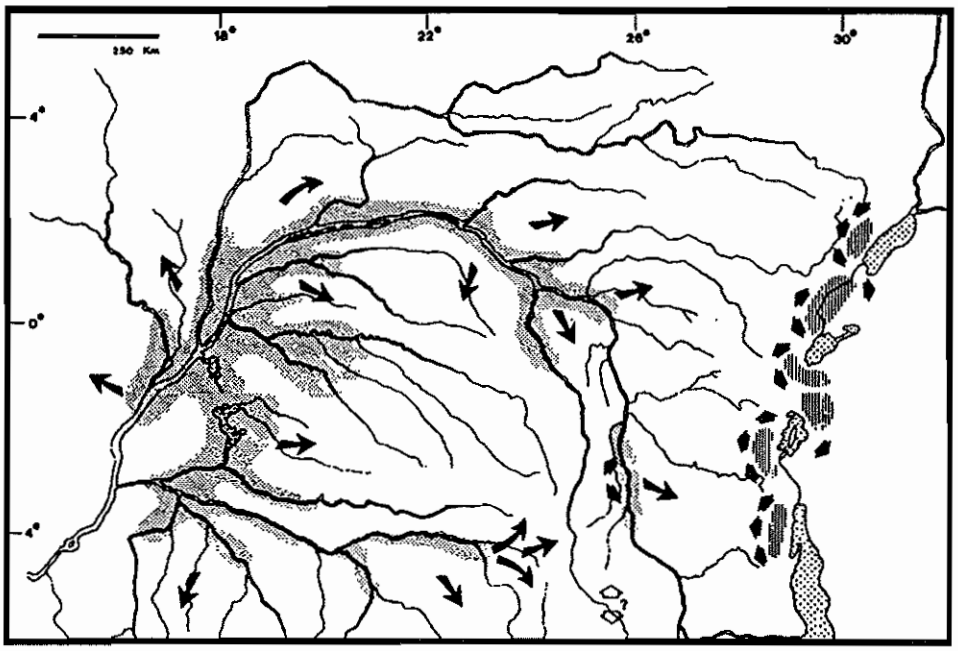


Figure 1.4: Pleistocene forest refugia. The top panel (from Hamilton, 1988) shows the traditional locations of the major African centers of endemism. The bottom panel shows Colyn's (1987) reinterpretation of the "East Zaïre Core Area" as a series of interconnected fluvial refugia, separate from the montane refugia of the Western Rift.

be negatively correlated with species diversity, since colonization rates of distant islands are generally low. In East African forests, primate and mammalian diversity decline predictably with distance from postulated refugia (Hamilton, 1976; Struhsaker, 1981; Rodgers, Owen and Homewood, 1982; Hamilton, 1988). Species-area relationships for primates imply that inter-forest dispersal may have been extremely limited among ancient refugia, but may not be significantly limited among extant forests (Chapman, 1983).

Pan troglodytes is a forest taxon in the sense that its current (and historical) distribution does not extend far beyond the limits of African forest (Kortlandt, 1972; Kortlandt, 1983; Teleki, 1989). Chimpanzees are adapted behaviorally and morphologically to life in a forest environment (Baldwin *et. al.*, 1981; Goodall, 1986; Wrangham, 1986; Fruth and Hohmann, 1994a; Wrangham *et. al.*, 1996). Chimpanzees can and do live in areas of high aridity (McGrew, Baldwin and Tutin, 1981; Kortlandt, 1983; Moore, 1992a; Sept, 1992). However, they cannot survive in the complete absence of forest habitat, probably due to dependence on forest foods and standing water (Kortlandt, 1983; Wrangham *et. al.*, 1991; Wrangham *et. al.*, 1993). No evidence has been found that chimpanzees have ever inhabited areas higher than approximately 15° north latitude, or lower than approximately 10° south latitude in Africa.

To the extent that chimpanzees have been restricted to forest habitats, their historical distribution should have tracked changes in the distribution of forest cover, as described above. During warm, wet phases (e.g. 120 kya, 10 kya), chimpanzees should have been widespread, distributed approximately throughout their current range. The onset of glacial conditions should have led to the vicariance of chimpanzee habitat. During periods of maximal aridity, chimpanzees should have been restricted to areas in and close to forest refugia . At 22 kya for example, chimpanzees should have been restricted to fluvial forest refugia in central Zaïre, and to medium-altitude montane forests in Uganda, Rwanda, Burundi, Tanzania and eastern Zaïre. As forests re-expanded during post-glacial climatic

amelioration (approximately 12.5 kya), chimpanzees should have rapidly dispersed out of these refugia into increasingly more distant areas. An analogous series of events should have occurred during each of the similar glacial/interglacial cycles which have characterized the late Pleistocene in Africa.

Genetic predictions

The cyclic contraction and expansion of the chimpanzee population in East Africa predicts that the subspecies experienced one or several bottlenecks in population size during its recent history. Bottlenecks reduce genetic variability in the populations which experience them (Nei, Maruyama and Chakraborty, 1975; Chakraborty and Nei, 1977; Watterson, 1984; Nei, 1987). Furthermore, the effects of a bottleneck persist over many generations following population re-expansion (Nei, Maruyama and Chakraborty, 1975). Bottleneck effects have been extensively studied in a variety of mammalian species (Bonnell and Selander, 1974; Nei and Graur, 1984; Sage and Wolff, 1986; O'Brien *et. al.*, 1987; O'Brien *et. al.*, 1989; McClenaghan, Berger and Truesdale, 1990; Packer *et. al.*, 1991; Sherwin *et. al.*, 1991; Wayne *et. al.*, 1991), and have been central in explaining low levels of mitochondrial genetic variability in humans (Cann, Stoneking and Wilson, 1987; Vigilant *et. al.*, 1991; Harpending *et. al.*, 1993; Sherry *et. al.*, 1994; Rogers, 1995a; but see Takahata, 1993). Nei and Graur (1984) and Sage and Wolf (1986) have interpreted low levels of genetic variability across a range of species as bottlenecks caused specifically by glacially-induced population size contractions.

Pleistocene refuge theory predicts that localized populations of eastern chimpanzees should have experienced different degrees and kinds of bottleneck effects. Populations inhabiting refugia should have been buffered from drastic size reductions. Populations currently inhabiting locations corresponding to Pleistocene refugia should therefore harbor greater genetic diversities than non-refuge populations. Moreover, the genetic diversity of a population should be a negative function of its distance from a refuge. The refuge model

predicts that non-refuge forests were colonized by animals dispersing out of refugia subsequent to global warming. Founder effects (Mayr, 1954) would occur under just such conditions, when small colonizing populations become isolated from parent populations. Because founder populations may carry only a subset of the alleles present in the parent population, they are likely to suffer reductions in genetic diversity (Mayr, 1954; Mayr, 1963; Nei, Maruyama and Chakraborty, 1975; Nei, 1987). Such reductions have been documented for several mammal populations (Wildt *et. al.*, 1987; Gilbert *et. al.*, 1990; McClenaghan, Berger and Truesdale, 1990; Wayne *et. al.*, 1991), and have been used to explain low genetic variability in humans (Flint *et. al.*, 1989; Tishkoff *et. al.*, 1996). Chimpanzee populations inhabiting the most distant forests should have been colonized though a series of sequential founder events. This "stepping-stone" model (Kimura and Weiss, 1964) would predict an incremental decline of genetic variability with distance from a source population. The predicted trend would be analogous to the documented decline of species diversity with distance from refugia, as described in the preceding section.

Gene flow among forests may also have been restricted subsequent to their colonization. If so, the eastern chimpanzee population should show some degree of population subdivisioning. Conspicuous biogeographic barriers (e.g. rivers, forest gaps) may correspond to sharp genetic divisions within the subspecies (Avise *et. al.*, 1987; Avise, 1994). Alternatively, populations may be differentiated only partially, or not at all (panmictic). The degree of population subdivisioning may be described within the framework of an "island model" of genetic differentiation (Wright, 1931; Wright, 1943). Under such a model, a population's genetic diversity is a function of its effective size and the amount of migration which it experiences (Wright, 1943; Nei, 1987). Effective population size, a measure of population genetic variability, is dependent on a variety of demographic and reproductive parameters (Wright, 1931; Kimura and Crow, 1963; Lande and Barrowclough, 1987; Avise, 1994). Although effective population sizes are generally

much lower than actual (census) sizes (Harris and Allendorf, 1989; Avise, 1994), the two quantities are correlated. An island model would therefore predict that the genetic diversity of a chimpanzee population should be positively correlated with the area of the forest which it occupies (assuming that the carrying capacities of all forests are similar). Because of the diversifying effects of migration, the same model would predict a positive correlation between diversity and proximity to other populations. Genetic studies of insular populations have indeed revealed such trends (Baker *et. al.*, 1990a; Gilbert *et. al.*, 1990; Inoue and Kawahara, 1990; Edwards, 1993; Seutin *et. al.*, 1993; Seutin *et. al.*, 1994; Westerbergh and Saura, 1994).

The model described above is fundamentally "dispersalist" in its formulation. The aforementioned processes of forest colonization involve "range expansion" and "jump dispersal" over biogeographic boundaries (Myers and Giller, 1988). However, the general importance of dispersal as a biogeographic process has been debated and most likely overestimated (Brundin, 1988; Myers and Giller, 1988). In this light, it is important to note that the genetic predictions described above would not change under an alternative vicariance model (Croizat, 1958; Croizat, 1964). Under such a model, chimpanzee populations would have become separated as forests fragmented during the onset of globally cool and dry conditions. Peripheral forests would have become fragmented first and most, causing severe bottlenecks and founder effects in the populations inhabiting them. Refuge forests would have become fragmented last and least, buffering populations within them against drastic size reductions. The genetic consequences of vicariance would thus be of the same nature and pattern as those hypothesized under the dispersalist scenario. The timing of these events would, however, coincide not with periods of rapid climatic warming, but with the more gradual onset of globally cool and dry conditions.

Hypotheses and structure of the dissertation

The predictive framework outlined above can be distilled into two biogeographic hypotheses, each making several genetic predictions:

1) *P. t. schweinfurthii* has experienced a recent demographic bottleneck. This hypothesis makes three genetic predictions: a) levels of genetic variability within eastern chimpanzees should be low (*i.e.* comparable to other taxa for which bottlenecks have occurred); b) because of the rapidity of transitions between cool, dry periods and warm, wet periods, the expansion should have been sudden, occurring over a time scale consistent with that over which forests re-expanded (2-3 thousand years); c) the timing of the expansion after the bottleneck should correspond to a documented warming event (e.g. 130 kya, 12.5 kya), most likely the 12.5 kya event.

2) *P. t. schweinfurthii* was restricted to Pleistocene forest refugia prior to expansion. This hypothesis makes three genetic predictions: a) the genetic diversities of refuge populations should be greater than those of non-refuge populations; b) population genetic diversity should decline incrementally with distance from a refuge; c) population genetic diversity should correlate positively with the size of the forest which the population occupies, and negatively with its distance from other populations (under an island model).

Chapter 2 describes the geographic sampling strategy adopted and the methodology used to collect the genetic data. Chapter 3 addresses Hypothesis 1a directly, by providing a descriptive account of eastern chimpanzee genetic variability. Chapter 3 also addresses Hypotheses 2a, 2b and 2c by examining the genetic variabilities of individual populations in light of the critical parameters of the biogeographic model. Chapter 4 analyzes the genetic data using "mismatch distribution" analysis (Rogers and Harpending, 1992). This statistical method, specifically designed to detect sudden demographic expansions,

addresses Hypotheses 1b and 1c directly. Chapter 5 addresses both main hypotheses using various techniques of phylogenetic reconstruction. Chapter 6 presents analyses unrelated to the biogeographic hypotheses described above. It tests hypotheses about the relationship between genetics and social behavior, which are explained at the beginning of the chapter. Conclusions and interpretations are presented for each chapter separately, and are synthesized in Chapter 7.

Chapter 2: Materials and Methods

Sample collection

The locations from which chimpanzees were sampled are shown in Figure 2.1. Sampling locations were chosen to represent a broad range of chimpanzee habitats, from insular montane forest on the eastern extreme of the subspecies range to expansive lowland rainforest in Zaïre. Detailed descriptions of Ugandan forests, which show remarkably complex ecological variation, may be found in Howard, 1991. Locations were also chosen to allow comparisons between populations on different sides of conspicuous biogeographic barriers (riverine and lacustrine systems). Chimpanzees were sampled from a total of 19 locations, representing 8 insular forests in Uganda, Rwanda and Tanzania, and 5 locations within the lowland Zaïrian forest block. In addition, 26 chimpanzees were sampled from a captive population housed at Entebbe Zoo, Uganda. The geographic origins of these chimpanzees are unknown. Names of sampling locations are given in Table 2.1, along with their longitudinal and latitudinal coordinates and numbers of samples used from each location. Abbreviated location names listed in Table 2.1 will be used throughout this study. A catalog of all samples used is given in Appendix 1.

Genetic material was collected non-invasively in the form of shed hair. Hair has proven a valuable source of genetic material for studies of wild chimpanzee populations, and for chimpanzee conservation (Moore, 1992b; Morin and Woodruff, 1992; Morin *et. al.*, 1994a). Chimpanzees typically construct between 1.1 and 2.1 nests per 24-hour period (Isabirye-Basuta, 1991). Although chimpanzees occasionally share or reuse nests, they typically construct a single new nest each night (Ghiglieri, 1984). Hairs collected from a nest are therefore likely to have come from a single individual. Mother-infant pairs, however, typically share a nest.

Collections were made between 6/91 and 11/93. Forests were searched for

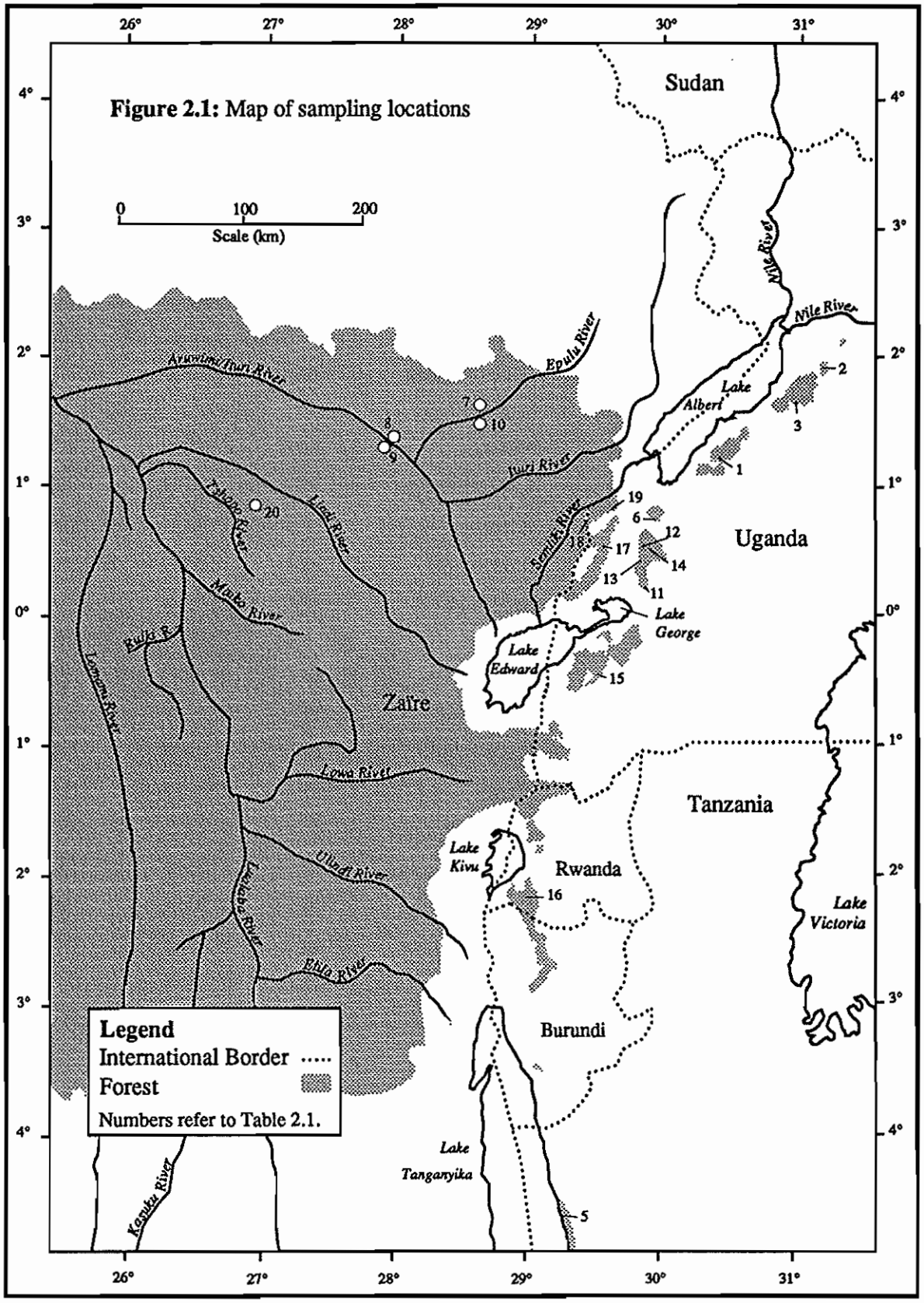


Figure 2.1: Map of sampling locations
Sudan
Uganda
Zaire
Tanzania
Rwanda
Burundi
Lake Albert
Lake Edward
Lake George
Lake Kivu
Lake Victoria
Lake Tanganyika
Nile River
Aruwimi/Ituri River
Epulu River
Ituri River
Semliki River
Lindi River
Tshopo River
Maiko River
Lowa River
Ulindi River
Elila River
Kamba River
Scale (km)
0
100
200
Legend
International Border
Forest
Numbers refer to Table 2.1.

Table 2.1: Description of sampling locations

	Abbrev.	*Forest*	*Location**	*Longitude***	*Latitude***	*Area (km^2)†*
1	BAMA	Bugoma	Mwela Sawmill	30° 59' E	1° 18' N	365
2	BOPI	Budongo	Pabidi	31° 44' E	1° 55' N	793
3	BOSO	Budongo	Sonso Sawmill	31° 33' E	1° 44' N	793
4	EEZO		Entebbe Zoo			
5	GEKA	Gombe	Kasekela	29° 36' E	4° 41' S	32
6	IARA	Itwara	Rwebikuya	30° 25' E	0° 47' N	87
7	IIAA	Ituri	Afarama	28° 32' E	1° 33' N	470,000
8	IIAE	Ituri	Avakubi NE	27° 34' E	1° 20' N	470,000
9	IIAW	Ituri	Avakubi SW	27° 33' E	1° 19' N	470,000
10	IILA	Ituri	Lenda	28° 38' E	1° 19' N	470,000
11	KEDN	Kibale	Dura Station	30° 19' E	0° 12' N	560
12	KEKA	Kibale	Kanyawara	30° 21' E	0° 34' N	560
13	KEKU	Kibale	Kanyanchu	30° 20' E	0° 28' N	560
14	KENO	Kibale	Ngogo	30° 25' E	0° 30' N	560
15	KUSL	Kalinzu	Kalinzu Sawmill	30° 03' E	0° 28' S	580
16	NESN	Nyungwe	Nyungwe Station	29° 30' E	2° 25' S	1140
17	RIKA	Rwenzori	Katebwa	30° 06' E	0° 32' N	996
18	SIMU	Semliki	Mbume-Busaru	30° 02' E	0° 43' N	212
19	SINI	Semliki	Ntandi	30° 09' E	0° 48' N	212
20	TOBA	Tshopo	Bafwabalinga	27° 04' E	0° 51' N	470,000

* See Figure 2.1; sampling locations are shown by number on this map.

** Latitudes and longitudes are measured to the nearest minute.

† Estimates of present-day areas of Ugandan, Tanzanian and Rwandan forests taken from Weber (1987) and Howard (1991); area estimate of Zaïrian forest (470,000) taken from Teleki (1989)

chimpanzee nests, which were removed from trees and searched for shed hairs. Searches were not systematic (*i.e.* transects were not used). Rather, local informants were employed to help concentrate searches in areas where chimpanzees were reputed to frequent. In most cases, information about the locations of community boundaries was unavailable. Nests from any single location may therefore represent individuals from multiple communities. The approximate area of forest searched varied across sampling locations, but was typically on the order of five km^2.

Each nest was assigned a 6-character identification code consisting of a four-letter forest location abbreviation (see Table 2.1) and a two-digit number (e.g. KEKA08). The following information was recorded for each nest: date of collection, height of nest (in meters), relative nest age (see below), and identity of the individual (if known). Nests were also assigned a single-letter "cluster designation." Nests were defined as in the same spatial cluster if they were within visual range of each other from the forest floor. All nests in a cluster were assigned the same single-letter code. Nests were assigned relative ages using a 5-point system, defined by the following criteria:

Age	*Nest characteristics*
1 (youngest)	Green leaves only
2	Mixture of green and brown leaves
3	Brown leaves only
4	Brown leaves, and < 50% of nest missing
5 (oldest)	Brown leaves, and >50% of nest missing

During collection, caution was exercised to minimize contamination of samples with foreign DNA. Hairs were handled exclusively with clean forceps. Hairs recovered from nests were sealed immediately in paper "stamp collector's" envelopes, which were carefully labeled. Collection envelopes were placed in airtight plastic envelopes packed with silica gel dessicant (Fisher Scientific, Fair Lawn NJ, catalog #S160-500). Hairs were stored dry

in the field at ambient temperature, and at -20° after transport to the United States.

A catalog of all samples used in this study (samples which yielded DNA sequences; see below) is given in appendix 1. In all populations except for KEKA, GEKA and EEZO, identities of individual chimpanzees were not known. Because nests were sampled randomly in most populations, individuals may have been sampled more than once. Hair from nests yielding different DNA sequences (see below) can be unambiguously assigned to different chimpanzee individuals. Hairs from nests yielding identical DNA sequences may have been shed by the same individual, or may represent shared, identical sequences from different individuals. The probability that identical DNA sequences resulted from double sampling of a single individual would depend on the number of chimpanzees in the community, the number of nests sampled, the ranging behaviors of individual chimpanzees, and the frequency of the sequence within the area. These parameters are unknown for any of the locations sampled. Possible biases associated with double sampling will therefore be discussed separately for each of the analyses performed.

Genetic analysis

The DNA region chosen for study was a hypervariable segment of the mitochondrial control region, also known as d-loop (Kocher and Wilson, 1991). D-loop is a non-coding region which is instrumental in the replication of the mitochondrial genome (Anderson *et. al.*, 1981). The control region can be further subdivided into two "hypervariable" regions, the first spanning bases 101-300 and the second spanning bases 601-900, initially identified by comparisons among human sequences (Kocher and Wilson, 1991). The first hypervariable region is the most quickly-evolving region in primate mitochondrial genomes, accounting for approximately 45% of variable positions in the human control region and 39% of variable positions in the chimpanzee control region (Kocher and Wilson, 1991). This region is therefore ideal for investigating evolutionary

relationships among closely-related taxa. The present study examines a 368 bp segment of the d-loop region (corresponding to Anderson reference sequence coordinates 16042 - 16410), which includes the first hypervariable region.

D-loop was chosen for study on several grounds. First, mitochondrial DNA does not recombine. This property simplifies laboratory analyses by obviating the need for cloning or other procedures for separating alleles. Second, mitochondrial genes are maternally inherited. Chimpanzees are a female-dispersing species in that females typically leave their natal communities at adolescence, whereas males typically remain (Pusey, 1980; Goodall, 1986). Mitochondrial genes should therefore be especially sensitive to dispersal barriers in this species. Third, mitochondrial genes evolve quickly (relative to nuclear genes). Hypervariable region 1 was selected for its especially high rate of evolution, because of the anticipated close evolutionary relationships of the taxa being studied. Finally, hypervariable region 1 has been well-characterized in humans and chimpanzees. This region has been used to support the "mitochondrial Eve" hypothesis of modern human origins (Cann, Stoneking and Wilson, 1987; Vigilant *et. al.*, 1989), and to discern patterns of chimpanzee phylogeography across Africa (Morin *et. al.*, 1994a). Data from the present study are directly comparable to these previously-published results.

DNA extraction

DNA was extracted from hair follicles using a modification of the methods described in Walsh *et. al.* (1991), which capitalize on the chelating properties of Chelex 100 resin (Bio-Rad, Richmond CA, catalog #143-2832). Chelex 100 binds polyvalent ions in solution; metal ions that normally catalyze DNA breakdown at high temperatures are rendered inactive by Chelex. Boiling can therefore be used to extract undamaged DNA from tissues of interest. The original Walsh *et. al.* protocol for hair follicles proved, upon preliminary trials, to yield sub-optimal quantities of DNA, especially for hairs that were not fresh. The protocol was therefore modified to incorporate enzymatic digestion of the hair

follicle prior to DNA extraction.

Hairs were removed from collection envelopes using flame-sterilized forceps. Hairs were examined for the presence of a follicle. Only hairs with follicles were chosen for extraction. From 1 to 3 hairs were used for each extraction. White, bulbous follicles (usually from fresh nests) were used alone; small or partially-degraded follicles were used in pairs; severely degraded follicles (usually from old nests) were used in triplets. Parallel negative controls (no hair added) were performed simultaneously with all extractions.

The follicle-containing end of the hair was washed briefly in 100% ethanol, and then rinsed thoroughly in distilled water. The proximal 1-2mm of hair shaft containing the follicle was severed using a flame-sterilized razor blade. The follicle was transferred to a 1.5 ml microfuge tube containing 250 µl of 5% Chelex 100 in distilled water. Seven µl of 1.0 M DTT and 1.5 µl of 20 mg/ml Proteinase K were added. The tube was incubated in a 56° heat block for between two and 12 hours.

After incubation, the tube was vortexed at high speed for 10 seconds and placed in a boiling water bath for 8 minutes. The tube was then vortexed for another 10 seconds and centrifuged at 14,000 g for 3 minutes in an Eppendorf microcentrifuge (model 5415 C). Two hundred µl of supernatant was transferred to a Microcon 30 tube (Amicon, Beverly MA, catalog #42410), without disturbing the Chelex resin. The 200 µl of ddH_20 was added to the original tube containing Chelex. The tube was vortexed for 10 seconds and centrifuged again at 14,000 g for 3 minutes. The 200 µl of supernatant was transferred to the Microcon tube containing the first supernatant. 100 µl of ddH_20 was added to the microcon tube, which was centrifuged for 8 minutes at 8,200 g. The retentate (approximately 10 µl) was collected and resuspended to 50 µl in ddH_20. Samples were used directly for PCR, or stored frozen at -20° C.

PCR

The polymerase chain reaction (PCR) is an *in-vitro*, enzymatic process used for amplifying small quantities of DNA target sequences (Saiki *et. al.*, 1985). PCR is facilitated by synthetic, sequence-specific oligonucleotide primers, designed to hybridize to sequences flanking the region of interest, which serve as initiation sites for sequence extension by *Taq* DNA polymerase. Typical reaction conditions and applications of the PCR technique are well-documented (e.g. Erlich, 1989; Innis *et. al.*, 1990).

As an assay for the presence of amplifiable DNA, a diagnostic PCR was performed on all samples. This PCR targeted a small (215bp) segment of the mitochondrial cytochrome oxidase subunit II gene (COII). COII is a protein-coding gene which is highly conserved among primates (Kocher and Wilson, 1991), evolving at rates approximately 1/10 those of the control region (Ruvolo *et. al.*, 1994). The short length of the fragment being amplified and the conserved nature of the gene sequence make this PCR a robust diagnostic for the presence of primate DNA. Ruvolo *et. al.* (1991) have demonstrated that primers designed to amplify this fragment hybridize to the appropriate gene regions in a wide taxonomic range of primates.

PCR was performed in a 12.5 µl volume with 5 µl of template using primers C187 (5'-TCAGACGCTCAGGAAATAGAAA-3') and D402 (5'-ACTCCTTGACGTTGACAA TCGA-3'); (Ruvolo *et. al.*, 1991; Disotell, 1994). Reaction conditions were as follows: 10mM Tris-HCl pH 8.3, 50mM KCl, 0.75 mM $MgCl_2$, 0.1 mM dNTPs, 2.5 pmol each of primers C187 and D402, and 0.3125 units of Taq DNA polymerase (Perkin Elmer, Branchburg NJ, catalog #N808-0106). PCRs were run on a Perkin Elmer model 480 thermocycler set to the following cycle parameters: 95° for 1 min., 50° for 1 min., 72° for 1 min., for 30 cycles followed by an indefinite soak at 4°. 5 µl of product was run on a 1.5% agarose gel impregnated with ethidium bromide in TAE buffer (0.04 M Tris-acetate, pH 8, 0.02 M EDTA) for 10 minutes at 200 volts . The gel was photographed under a 320

nm ultraviolet light, and the presence or absence of a band was scored for each sample.

A strong, monotonic decrease was observed in the proportion of samples that yielded visible product as a function of nest age (Figure 2.2). This observation suggests that DNA becomes degraded over time in nest hairs. DNA from old hairs is thus "ancient," in that it is scarce, fragmented, or otherwise resistant to amplification (Pääbo, Gifford and Wilson, 1988; Pääbo, 1990). This conclusion is strengthened by the fact that all attempts to amplify segments of the COII gene larger than approximately 500 bp failed for hair-extracted samples of relative ages greater than 2. Primers used in these attempts nevertheless worked consistently on chimpanzee DNA extracted from whole blood or tissue, as well as on a range of other primate taxa (Ruvolo *et. al.*, 1991; Disotell, 1994).

Because of the degraded nature of the DNA, a two-stage, "nested" PCR strategy was used to generate quantities of template suitable for sequencing. The target sequence (3' end of the mitochondrial control region) was amplified in two separate stages: one initial stage in which the entire first hypervariable region was amplified, and one secondary amplification, in which the initial product was re-amplified in two overlapping fragments (Figure 2.3). Primers were designed based on published human and chimpanzee sequences (Anderson *et. al.*, 1981; Kocher and Wilson, 1991; Horai *et. al.*, 1992; Morin *et. al.*, 1994a). Four primers were used for nested PCR and sequencing (L-16041: 5'-CTCTGTTCTTTCATGGGGAAGC-3'; L-16111: 5'-ATTTCGTACATTACTGCCAGCC-3'; H-16286: 5'-GGATGGATTTGACTGTAATGTGC-3' ; H-16411: 5'-TGTGCGGGAT ATTGATTTCACG-3') and a fifth primer was used for sequencing only (H-16128: GGTAGTTGAGTGATTATAGTACTG). Primer names indicate the strand to which the primer hybridizes (L = light, H = heavy), and numbers indicate locations of the 3' termini, relative to the human reference sequence (Anderson *et. al.*, 1981). Negative controls (Chelex extractions without hair follicles) were run with all PCRs to test for possible contamination by exogenous DNA.

Initial PCR amplifications were performed in a 12.5 µl volume using 2.5 µl of

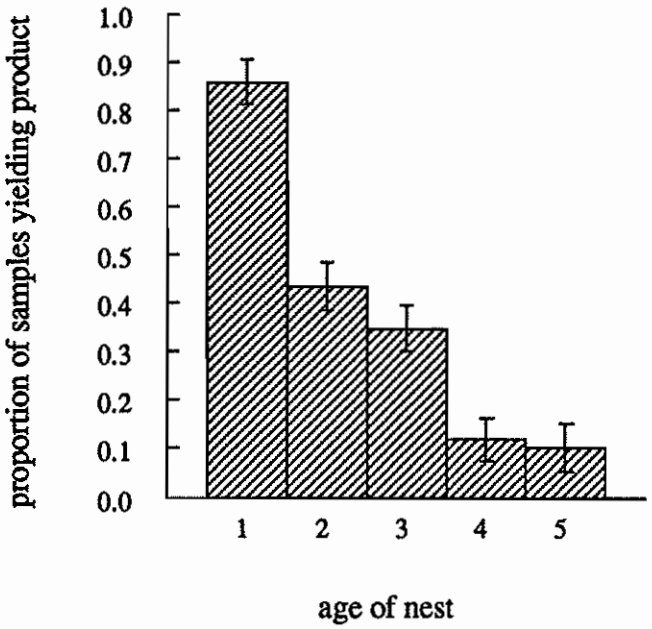


Figure 2.2: Success of DNA amplification by PCR as a function of sample age. DNA was extracted from hair collected from 352 nests of relative ages 1 (youngest) to 5 (oldest). PCR was performed on a 215 bp segment of the COII gene (see text for full explanation). Presence of visible product decreased significantly with nest age (Likelihood Ratio Chi-Square = 91.8; $p << 0.001$; error bars represent standard error of the mean).

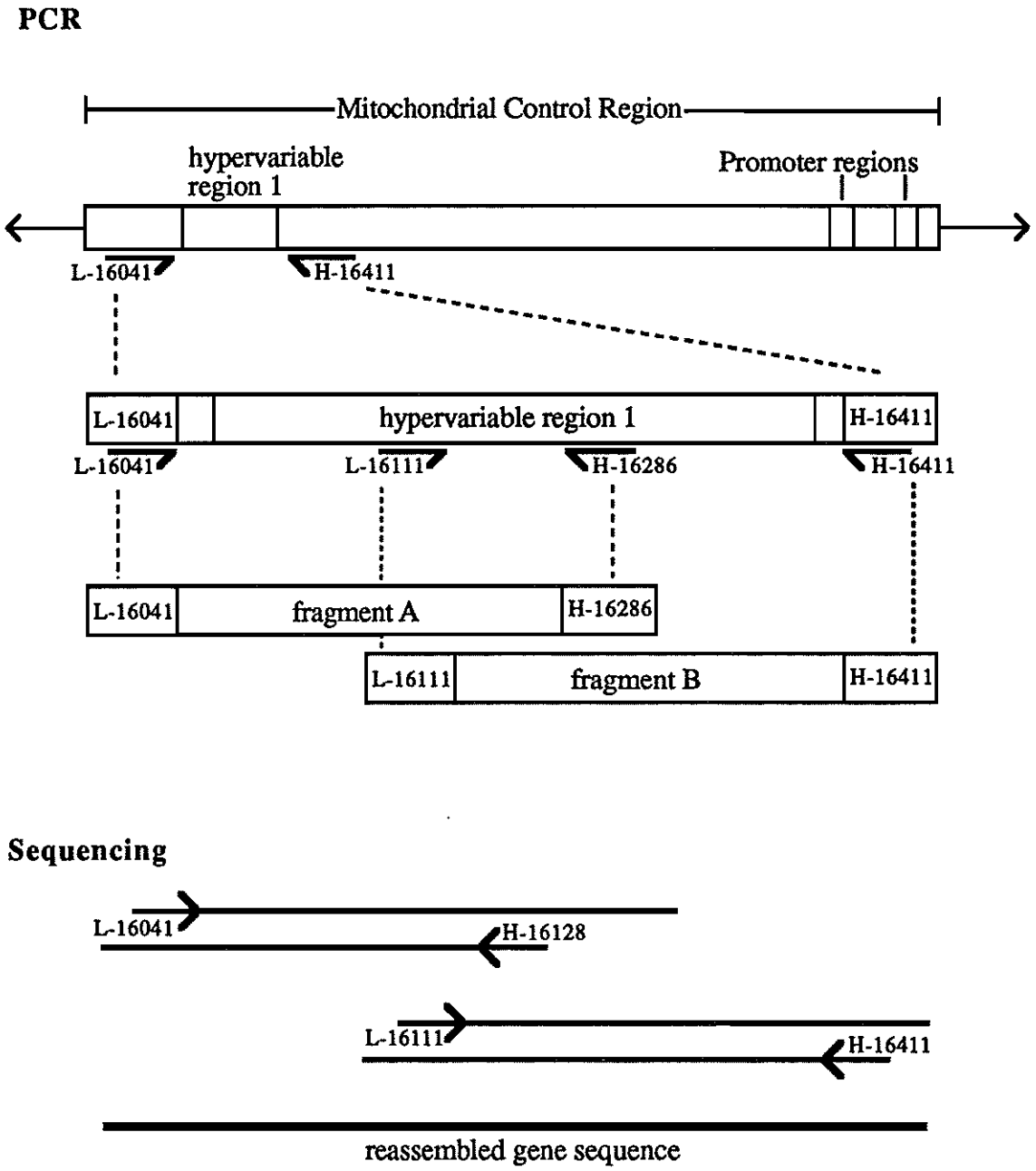


Figure 2.3: Schematic representation of PCR and sequencing strategies used in the study. Hypervariable region 1 of d-loop was amplified from genomic DNA isolated from hair. The resulting product was gel-purified and reamplified in 2 overlapping fragments. Each fragment was sequenced twice using an ABI model 373 automated DNA sequencer: fragment A with primers L-16041 and H-16128, and fragment B with primers L-16111 and H-16411. The gene sequence was reassembled by computer from the four resulting chromatograms (see text for full explanation).

genomic DNA extracted from hair. Reaction specifications were as follows:10mM Tris-HCl pH 8.3, 50mM KCl, 2.0 mM $MgCl_2$, 0.2 mM dNTPs, 6.25 pmol each of primers L-16041 and H-16411, and 0.3125 units of Taq DNA polymerase. PCRs were run in a Perkin-Elmer model 9600 thermocycler set to the following cycling parameters: 95° for 30 sec., 53° for 30 sec., 72° for 30 sec., for 45 cycles followed by an indefinite soak at 4°. The thermocycler was preheated to 95° prior to loading to reduce non-specific priming and extension during initial ramping.

Products were lyophylized and resuspended in 1x loading buffer consisting of 7% (wt/vol) sucrose and 0.05% Xylene cyanol FF in water (Sambrook, Fritsch and Maniatis, 1989). Samples were loaded onto a 1.0% low-melt agarose gel (SeaPlaque GTG agarose, American Bioanalytical, Natick MA, catalog #AB197) impregnated with ethidium bromide and run for approximately 30 minutes at 100 volts in TAE buffer. The gel was then photographed under 320nm ultraviolet light. Fresh hairs and positive control DNA (isolated from *P. t. verus* placental tissue) always yielded high-intensity bands in the expected locations. Old hairs, however, yielded weak bands or no visible bands at all.

Single plugs were removed from the gel from the center of the band region in each lane, using a sterile pasteur pipette. Plugs were removed from all lanes, regardless of whether a band was visible, and were transferred to 1.5 ml microfuge tubes. Care was taken to minimize exposure of the gel and sample to UV light during this procedure. Tubes were centrifuged briefly to concentrate the plug and were placed in a 68° water bath for 5 minutes to melt the agarose. Tubes were then transferred to a 37° water bath for 10 minutes. One µl of agarase (5 u/µl; Sigma, St. Louis MO, catalog #A-6306) was added to each tube and mixed thoroughly. Samples were incubated at 37° for 1-10 hours, after which they remained liquid at room temperature. Plugs from bright bands were diluted 50 fold with water prior to reamplification. Plugs from weak, but visible, bands were diluted 20 fold. Plugs from lanes where no band was visible were left undiluted. Digested plugs

were used directly for second-round PCR, or were stored at -20°.

The original PCR product was reamplified in two separate reactions, producing two overlapping fragments. Primers L-16041 and H-16286 were used to generate the first fragment, and primers L-16111 and H-16411 were used to generate the second. Reamplifications were run in a volume of 50 µl, with 2.0 µl of template from the initial amplification. The reaction conditions for both reamplifications were identical: 10mM Tris-HCl pH 8.3, 50mM KCl, 1.0 mM $MgCl_2$, 0.2 mM dNTPs. Reactions were run in a Perkin Elmer model 9600 thermocycler set to the following cycling parameters: 95° for 30 sec., 55° for 30 sec., 72° for 30 sec., for 40 cycles, followed by an indefinite soak at 4°. The thermocycler was preheated to 95° prior to sample loading to reduce non-specific priming and extension during initial ramping.

Products were lyophylized and resuspended in 8.5 µl of 1x loading buffer (7% wt/vol sucrose and 0.05% Xylene cyanol FF). The entire reaction was loaded onto a 1% SeaPlaque GTG low-melt agarose gel (FMC BioProducts, Rockland, ME., catalog # 50112) impregnated with ethidium bromide and run for approximately 20 minutes at 100 volts in TAE buffer. The gel was transferred to an ultraviolet light box and photographed under 320 nm uv light. Visible bands were excised using a sterile razor blade, and transferred to 1.5 ml microfuge tubes. The bands were digested with agarase as described above, and were resuspended in water to 70 µl. DNA was used directly in automated sequencing (see below), or stored at -20°.

Extreme precautions were taken to avoid contamination during all stages of this protocol (Kwok, 1990). Because of the efficiency of PCR, very small amounts of extraneous template (as little as one molecule) can contaminate a sample. The potential for contamination is especially high in nested PCR because of the sensitivity of the procedure and the number of manipulations involved. Hair extractions and PCR set-up were performed in areas separate from those in which PCR products were handled. All reagents

were made from sterile solutions. All gel-running equipment was soaked thoroughly in 5% bleach before use. Gel casting trays and combs were soaked in bleach and exposed to ultraviolet light for 10 minutes prior to use. Appropriate negative controls were run during all steps. If contamination was detected in any negative control, the experiment was aborted and fresh preparations were made.

DNA sequencing

All sequencing was performed on an ABI model 373 automated DNA sequencer (Applied Biosystems, Foster City CA). Unlike traditional sequencing methods, which involve extensive template preparation and rely on hazardous radioisotopes or toxic chemicals (Kusukawa *et. al.*, 1990; Tahara, Kraus and Rosenberg, 1990; Medori, Tritschler and Gambetti, 1992; Douglas, Georgalis and Atchison, 1993), the ABI sequencing strategy utilizes termination nucleotides labeled with florescent dyes (Applied Biosystems, 1994a). The protocol employs a "cycle sequencing" approach in which a different dye is tagged to each of the four different termination nucleotides (Applied Biosystems, 1993). The product is then scanned by a laser during electrophoresis through a denaturing acrylamide gel. The resulting data are analyzed by computer and translated into a chromatogram, in which each base is represented by a line of different color (Figure 2.4). Although the signal peaks are read by the computer to produce a sequence, all chromatograms were visually inspected to monitor the accuracy of computer base-calling. Chromatograms are analogous to traditional sequencing audoradiograms, except that all four bases appear in the same lane. The advantages of this technique over traditional "manual" sequencing include ease of sample preparation, increased length of reads, increased numbers of samples that can be processed simultaneously, and increased accuracy (Applied Biosystems, 1993; Hyder *et. al.*, 1994).

Four initial sequencing reactions were performed for each sample (see Figure 2.3).

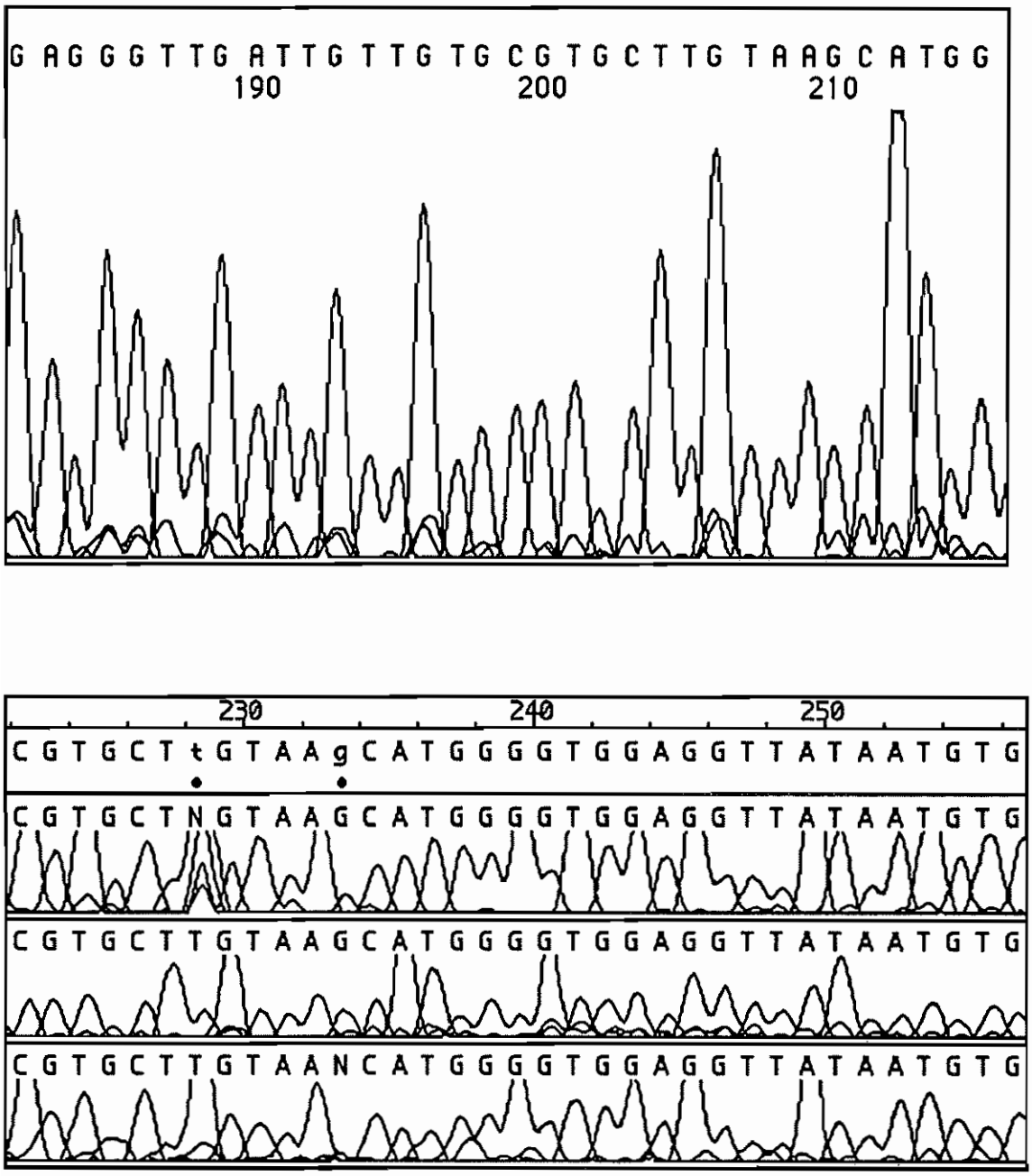


Figure 2.4: Chromatograms produced by an ABI model 373 automated DNA sequencing system. Top panel shows chromatogram output from a single d-loop sequence. Bottom panel shows three sequences aligned with the Auto Assembler program. Note resolution of ambiguities (indicated by "•") by repeated sequencing.

The fragment amplified with primers L-16041 and H-16286 was sequenced with primers L-16041 and H-16128. The fragment amplified with L-16111 and H-16411 was sequenced with primers L-16111 and H-16411. 5 µl of template was used in each reaction. Sequencing reactions were run in a volume of 20 µl, containing 9.5 µl of reaction premix and 3.2 pmol of sequencing primer. Sequencing was performed in a Perkin Elmer model 9600 thermocycler, following the specifications outlined by ABI (Applied Biosystems, 1993). The four resulting sequences were then assembled using software provided by ABI (Applied Biosystems, 1994b). Sequences were considered complete when confirmed by two independent reads. All ambiguities remaining after initial sequencing were resolved by re-sequencing with a different primer, or by altering template concentration (Hyder *et. al.*, 1994).

Sequences were aligned by hand. Since no insertions or deletions were observed within the chimpanzees sequenced, alignment was straightforward. Published sequences from Morin *et. al.* (1994a) were aligned with reference to the sequences generated in this study; gaps in the Morin sequences were treated as missing data in phylogenetic analyses.

Sequence reliability

The sensitivity of nested PCR to contamination, coupled with the lack of 3' to 5' exonuclease "proofreading" ability in *Taq* DNA polymerase, raises concern over the reliability of the sequences generated. This concern is heightened by the fact that the protocol described above employs *Taq* polymerase three times in succession (once for each round of nested PCR, and once during cycle sequencing). Estimates of *Taq* polymerase fidelity range from approximately 1.7×10^{-4} per nucleotide polymerized per cycle (Saiki *et. al.*, 1988) to 5×10^{-6} per nucleotide polymerized per cycle (Fucharoen *et. al.*, 1989). Fidelity rates vary considerably according to reaction specifications. Low fidelity PCR conditions have even been used for random mutagenesis of DNA sequences (Leung, Chen and Goeddel, 1989). However, other studies have documented high fidelity

rates (approximately 10^{-5}) when concentrations of dNTPs and MgCl2 have been kept low (Fucharoen *et. al.*, 1989; Goodenow *et. al.*, 1989).

The reactions specifications in the present study conform to high-fidelity conditions, as defined and documented in the literature. A given sample in this study will have been subjected to 85 cycles of PCR and 25 cycles of *Taq*-based cycle sequencing during analysis. Assuming a misincorporation rate of 10^{-5} per nucleotide per cycle, the expected fidelity per sample is $110(10^{-5}) = 2.2 \times 10^{-5}$ per nucleotide. For a sequence of 400 bp, the expected number of misincorporations is 8.8×10^{-3}. In a data set of 300 such sequences, the estimated total number of misincorporations is 2.64, or roughly three base pairs, which represents a total cumulative error of 0.002%. Thus, even in this "worst-case" scenario (the parameters used in the above calculations are high-end estimates), the expected error due to nucleotide misincorporation by *Taq* polymerase is negligible. Furthermore, the probability of misincorporation of the same base at the same position in each of 110 PCR cycles is astronomically low. Random misincorporation, a more likely scenario, would be invisible during sequencing; the effect of a single or a few miscopied template molecules would be swamped by the majority of correctly-copied ones. Ambiguous signals at a site were always re-sequenced using independent reactions.

Concern has also been expressed over the reliability of mitochondrial DNA sequences due to of the presence of similar, potentially cross-hybridizing sequences in the nuclear genome (Fukuda *et. al.*, 1985; Nomiyama *et. al.*, 1985; Irwin, Kocher and Wilson, 1991; Zischler *et. al.*, 1995). These sequences are ubiquitous in humans and primates, and provide evidence that exchange of genetic information between the two genomes has been a consistent feature of genomic evolution (Hu and Thilly, 1994; Stewart, 1995). Nuclear copies of mitochondrial genes can be long and contiguous (Kamimura *et. al.*, 1989), and can serve as alternate sites for annealing of PCR primers designed to hybridize to mtDNA (Irwin, Kocher and Wilson, 1991; Stewart, 1995). Inadvertent

amplification of a nuclear copy of a mitochondrial gene would confound phylogenetic analysis, since a separately-evolving nuclear sequence (unless very recently inserted) would not be homologous to the corresponding mitochondrial sequence.

Several lines of evidence suggest that such inadvertent amplification of nuclear DNA has not occurred in the present study. If a second, nuclear copy of the control region had been amplified along with the intended mitochondrial copy, then the duplicate sequencing reads used in this study would reveal nucleotide positions in which bases were ambiguous and unresolvable (Irwin, Kocher and Wilson, 1991). No sequences in this study showed such a pattern. However, an inserted nuclear copy of a mitochondrial sequence might have been amplified to the exclusion of the target sequence. Such a nuclear copy would likely be an outlier among its mitochondrial paralogs in that it would show unusual patterns of nucleotide substitution or conspicuously different evolutionary rates. Again, no such sequences were found in the present study. Finally, the probability that nuclear sequences were inadvertently generated is low because all directed attempts to amplify nuclear genes from field-collected samples resulted in failure. Several preliminary attempts were made to amplify short (100-200 bp) "microsatellite" regions from DNA extracted from hair (Jeffreys, Wilson and Thein, 1985; Ely *et. al.*, 1992; Morin *et. al.*, 1994a). None of these attempts was successful, perhaps because of the degraded nature of DNA extracted from hair, coupled with the low copy number of nuclear DNA per cell relative to mitochondrial DNA. The failure of these attempts argues against the possibility that nuclear copies of control region were inadvertently amplified and sequenced.

As a direct test of the reliability of the PCR amplification and sequencing strategy used in this study, ten samples were chosen for re-sequencing. Nested PCR from purified genomic DNA was performed twice on each sample as described above. The resulting PCR products were sequenced and compared. In all 10 cases, the second sequence exactly matched the first.

Chapter 3: Biogeographic predictors of genetic diversity and distance

Introduction

Little is known about the extent or apportionment of genetic diversity within any ape subspecies. Chimpanzees (*Pan troglodytes*), gorillas (*Gorilla gorilla*) and orang-utans (*Pongo pygmaeus*) have been separated into subspecies on the basis of geographic distribution (Schwartz, 1934; Groves, 1971; von Koenigswald, 1982) and morphological differences (Coolidge, 1929; Groves, 1970; Jacobshagen, 1974; Shea and Coolidge, 1988). In all three species, genetic distances among subspecies exist (Ferris, Wilson and Brown, 1981; Garner and Ryder, 1992; Morin *et. al.*, 1994a; Ruvolo *et. al.*, 1994). Chimpanzee subspecies correspond to clades on a phylogenetic tree (Morin *et. al.*, 1994a), suggesting that subspecies represent reproductively-isolated populations.

In contrast, humans do not show significant genetic differentiation at a racial level, despite an impressive geographic range (Lewontin, 1972; Latter, 1980; Cann, Stoneking and Wilson, 1987; Stoneking *et. al.*, 1990; Merriwether *et. al.*, 1991; Vigilant *et. al.*, 1991; Ward *et. al.*, 1991; Ruvolo *et. al.*, 1993). Data on mitochondrial variability within ape species suggest that humans may represent the low end of the diversity spectrum within the hominoids (Brown, 1980; Ferris, Wilson and Brown, 1981; Ruvolo *et. al.*, 1994). However, the existence of isolating reproductive barriers among ape subspecies would render comparisons between humans and apes at the species level misleading, since humans are not phylogeographically subdivided in a comparable way. Preliminary data by Morin *et. al.* (1994a) indicate that within-subspecies diversity in chimpanzees is lower than between-subspecies diversity, and is generally higher than that within humans. Unfortunately, the geographic origins of many of Morin *et al.*'s chimpanzees are unknown, and their DNA sequences are largely incomplete.

This study attempts to provide the first geographically systematic genetic survey of

an ape subspecies, the eastern subspecies of chimpanzee, *Pan troglodytes schweinfurthii.* Eastern chimpanzees range as far west as the Ubangi River in Zaïre and extend eastward to the montane forests of the western Rift (see Chapter 1). The northern, southern and western extremes of the subspecies distribution are demarcated by major rivers. The eastern extreme of the subspecies distribution is demarcated by the eastern limits of suitable forest habitat. Within this range, eastern chimpanzees inhabit two qualitatively different habitat types: expansive lowland rainforest in Zaïre, and smaller forest "islands" on the eastern borders of the subspecies range. In the large Zaïrian forest block, potential biogeographic barriers (rivers) transect the subspecies range. In the eastern region, montane forest islands are separated by gaps in forest cover, which may similarly restrict interpopulational gene flow.

Individual samples were collected from 19 geographically-localized populations across a representative range of eastern chimpanzee habitat types (see Chapter 2 and Appendix 1). Two hundred and fifty five complete DNA sequences were generated from these samples. These sequences, plus nineteen sequences taken from Morin *et. al.* (1994a), yielded 123 unique haplotypes with a total of 90 variable nucleotide positions (Appendix 2). Of these 123 haplotypes, 99 were population-specific, and 24 were shared among populations (Appendix 3).

The first section of this chapter provides a quantitative description of the levels and apportionment of genetic diversity within the eastern chimpanzee subspecies based on these data. The following sections attempt to explain why genetic diversity in eastern chimpanzees is apportioned as it is. Specifically, these sections test predictive hypotheses about chimpanzee genetic evolution which are derived from biogeography theory (see Chapter 1). Differences in genetic diversity among populations are explained with reference to a theoretical framework drawing upon: 1) the theory of insular biogeography (MacArthur and Wilson, 1967) and 2) reconstructions of East African Pleistocene forest

refugia (Hamilton, 1981; Grubb, 1982; Colyn, 1991). Genetic distances among populations are explained with reference to the current geographic distribution of chimpanzees in East Africa.

Genetic diversity

Genetic diversity measures were calculated based on the 123 haplotypes (24 shared among populations, 99 "private") described in Appendices 2 and 3. Simple computer programs were written to calculate each measure. Five measures of genetic diversity are included in the analyses which follow. Each measure has different properties which render it uniquely informative. The first measure is the well-known Shannon-Wiener Diversity Index, which is an "industry standard" in community ecology and biogeography (Peet, 1974; Brown, 1988). It is given by:

$$H' = -\sum_{i=1}^{n} p_i (\ln p_i)$$

where p_i represents the frequency of the *i*th element in a system. The Shannon-Wiener Index is useful for capturing information simultaneously about the number of elements in a system and their relative abundances. Although not frequently used in genetic analyses because of its unclear genetic meaning (Nei, 1987), it is easily adaptable to genetic data if p_i is taken to be the frequency of the *i*th allele at a locus (Lewontin, 1972).

Like the Shannon-Wiener Index, Nei's (1973) "gene diversity" assumes equidistance among all haplotypes. An unbiased estimate of gene diversity is calculated as:

$$\hat{h} = (n/(n-1))(1 - \sum_{i=1}^{m} x_i^2)$$

where x_i is the frequency of the *i*th allele at a locus of m alleles in a population sample of size n. It is commonly called "heterozygosity" in the literature, although this label belies its usefulness for haploid and polyploid genetic systems (Nei, 1973).

"Nucleotide diversity" is described by Nei (1987) as "heterozygosity at the nucleotide level," and is the same as average percent sequence difference among individuals. An unbiased estimated of nucleotide diversity is:

$$\hat{\pi} = (n/(n-1))(\sum_{ij} x_i x_j \pi_{ij})$$

where x_i and x_j are the frequencies of the *i*th and *j*th haplotypes in a population sample of size n, and π_{ij} is the proportion of different nucleotides (per site) between them. Unlike the previous two diversity measures, nucleotide diversity directly measures genetic variability at the DNA sequence level.

Nucleotide diversity can be thought of as the mean pairwise sequence difference among haplotypes within a population. Modal and maximum pairwise sequence difference among haplotypes are also included in the analysis. Modal pairwise sequence difference is necessarily correlated with nucleotide diversity, but is included because of its relevance to mismatch distribution analysis (Rogers and Harpending, 1992) described in Chapter 4. Maximum pairwise sequence difference is included because of its relevance to neutral coalescent theory (Tajima, 1983).

The latter three diversity measures estimate genetic diversity at the level of the DNA sequence. Interpopulational differences in these measures (barring sampling error) therefore result from the generation and fixation of point mutations, which is a relatively slow process (Li and Graur, 1991). In contrast, the Shannon-Wiener Index and gene diversity ignore molecular distances among haplotypes. Interpopulational differences in these measures will therefore accumulate from the drift and fixation of entire alleles. Since point mutation is largely removed from the equation, gene diversity and Shannon-Wiener diversity provide a more recent window into evolutionary history than do sequence-based measures.

Table 3.1 presents sample sizes and genetic diversity measures for each population. EEZO population is included in this table, although the geographic provenances of these

Table 3.1: Measures of genetic diversity for individual sampling locations

			Diversity Measure†				
	*Location**	*n***	*SW*	*GD*	*ND*	*MD*	*XD*
1	BAMA	13	1.29	0.64	0.85	0.00	2.65
2	BOPI	12	1.94	0.91	1.61	1.78	3.56
3	BOSO	15	1.97	0.89	1.55	2.04	2.62
4	EEZO	26	3.04	1.02	2.04	1.70	4.80
5	GEKA	19	2.31	0.94	2.02	2.29	3.73
6	IARA	12	1.94	0.91	1.55	1.19	3.85
7	IIAA	13	2.20	0.95	2.97	3.24	5.00
8	IIAE	12	1.23	0.67	1.21	0.00	2.96
9	IIAW	13	1.99	0.92	2.25	2.36	3.83
10	IILA	17	2.28	0.94	1.87	1.88	3.75
11	KEDN	14	1.30	0.73	1.97	2.93	3.51
12	KEKA	15	1.84	0.89	0.97	1.16	1.75
13	KEKU	14	1.77	0.82	2.22	3.22	3.80
14	KENO	13	1.48	0.81	2.06	0.88	4.42
15	KUSL	13	2.35	0.97	1.59	1.47	4.12
16	NESN	13	1.88	0.91	1.87	2.06	2.94
17	RIKA	13	2.35	0.97	2.70	3.24	5.30
18	SIMU	13	1.26	0.69	1.99	0.00	4.12
19	SINI	11	1.85	0.90	2.27	2.09	3.89
20	TOBA	10	1.83	0.91	1.84	1.81	3.32
Eastern Forests		171	3.96	0.98	2.04	1.91	5.47
Zaïrian Forests		65	3.54	0.98	2.16	1.93	4.97
Total		281	4.43	0.98	2.08	1.91	5.45

* For full description of sampling locations, see Table 2.1 and associated text.

** Sample sizes of DNA sequences

† Diversity measures are:

SW: Shannon-Wiener Index

GD: Nei's (1987) Gene Diversity

ND: Nei's (1987) Nucleotide Diversity (%)

MD: Modal pairwise sequence difference (%) for all pairs of individuals

XD: Maximum pairwise sequence difference (%) for all pairs of individuals

animals are unknown. Data are also presented for Zaïrian forests (IIAA, IIAE, IIAW, IILA and TOBA) and eastern forests (all other locations except EEZO) separately, and for the total combined population of eastern chimpanzees. Diversity measures vary widely among populations, with some qualitative consistencies. EEZO ranks consistently high, as might be expected if these zoo animals (confiscated from poachers and illegal animal traders) originated from geographically-disparate areas. Of the geographically-defined populations, IIAA (a Zaïrian forest) and RIKA (an eastern forest) appear at the upper end of the diversity spectrum by most measures. In most cases, BAMA is at the lower extreme. Examination of the diversity of the combined eastern samples and the combined Zaïrian samples reveals no consistent difference. Independent t-test comparisons of means revealed no statistically significant difference between the eastern and Zaïrian regions for any diversity measure.

To compare eastern chimpanzee diversity directly to human diversity, Vigilant's (1991) world sample of 135 human d-loop sequences was analyzed. Human sequences were aligned by hand to chimpanzee sequences. Human sequences were then edited to exclude all but the homologous 368 base pair region of d-loop sequenced in the present study. Nucleotide diversity (*i.e.* mean pairwise sequence difference) in humans was calculated as 2.62%, which is higher than the corresponding eastern chimpanzees estimate (2.08%). When all identical haplotypes within the eastern chimpanzee sample were collapsed (resulting sample size = 123 sequences), nucleotide diversity in eastern chimpanzees increased to 2.14% . Modal sequence difference within humans (2.72%) was similarly higher than the corresponding eastern chimpanzee estimate (1.91%). Finally, maximum pairwise sequence difference within humans (the parameter on which the "mitochondrial Eve" hypothesis of modern human origins is largely based), is also slightly higher in humans (5.97%) than in eastern chimpanzees (5.45%; see also Chapter 5).

Endemism

Although not a measure of diversity *per se*, endemism is a valuable concept for the analysis of biogeographic data. In biogeography, a taxon is endemic if it exists exclusively within a restricted geographic range (Major, 1988). This definition is clearly arbitrary, in that every taxon is endemic on some geographic scale. Nevertheless, the concept of endemism is specifically relevant to Pleistocene refuge theory, since refugia are often identified or defined partly on the basis of the large numbers of endemic taxa which they contain (Grubb, 1982; Haffer, 1982; Colyn, 1991; Colyn, Gautier-Hion and Verheyen, 1991). Taxa with cosmopolitan distributions are "eurytopic."

This study applies the concept of endemism to genetic data at the allelic level. A haplotype was defined as endemic if it appeared exclusively in one sampling location. If a haplotype appeared in more than one sampling location, it was defined as eurytopic. "Endemic" alleles are probably not truly endemic. Many likely occur in low-frequencies in populations in which they were not sampled. Similarly, eurytopic alleles are probably shared among more sampling locations than the data suggest. Endemism (the proportion of endemic haplotypes in a population) is therefore a sample-size-dependent diversity measure which emphasizes low-frequency alleles.

Table 3.2 presents numbers of endemic and eurytopic haplotypes for each population. Frequency distributions of endemic haplotypes are also presented in Table 3.2. Population frequencies of the 24 identified eurytopic haplotypes are presented in Table 3.3. Of the 123 haplotypes identified in the study, 99 (80%) were endemic. Endemism is clearly higher in Zaïrian forests than in eastern forests. Within Zaïrian forests, all haplotypes were endemic. In other words, although haplotypes were shared between Zaïrian and eastern locations, no haplotypes were shared among locations within Zaïre. In contrast, 73% of eastern haplotypes were shared among eastern sampling locations. To compare eastern and Zaïrian forests individually, endemism was quantified for each sampling location as the percent of haplotypes which were endemic to that area. Endemism

Table 3.2: Numbers of endemic and eurytopic and haplotypes identified in individual sampling locations

	*Location**	*Sample size (n)*	*Number of haplotypes*	*Number of endemic haplotypes***	*Number of eurytopic haplotypes†*	*Frequency distribution of endemic haplotypes‡*
1	BAMA	13	6	2	4	1,1
2	BOPI	12	8	4	4	1,1,1,1
3	BOSO	15	9	4	5	1,1,1,1
4	EEZO	26	22	11	11	1,1,1,1,1,1,1,1,1,1,1
5	GEKA	19	11	10	1	3,3,2,2,2,2,1,1,1,1
6	IARA	12	8	4	4	1,1,1,1
7	IIAA	13	10	10	0	3,2,1,1,1,1,1,1,1,1
8	IIAE	12	5	5	0	7,2,1,1,1
9	IIAW	13	7	6	1	2,2,2,1,1,1
10	IILA	17	11	10	1	3,3,2,1,1,1,1,1,1,1
11	KEDN	14	5	2	3	5,1
12	KEKA	15	6	1	5	1
13	KEKU	14	8	0	8	
14	KENO	13	5	1	4	3
15	KUSL	13	11	8	3	1,1,1,1,1,1,1,1
16	NESN	13	7	7	0	3,2,2,2,2,1,1
17	RIKA	13	11	3	8	2,1,1
18	SIMU	13	5	3	2	1,1,1
19	SINI	11	7	3	4	2,1,1
20	TOBA	10	7	6	1	3,2,1,1,1,1
Eastern Forests [a]		190	75	55	20	
Zaïrian Forests [a]		65	41	41	0	
Total		281	123	99	24	

* For full description of sampling locations, see Table 2.1 and associated text.

** Endemic haplotypes are defined as haplotypes which appear exclusively in one sampling location.

† Eurytopic haplotypes are defined as haplotypes which appear in more than one sampling location.

‡ Frequency distributions show absolute numbers of each distinct endemic haplotype found in each sampling location; for frequencies of eurytopic haplotypes, see Table 3.3.

[a] Numbers of endemic and eurytopic haplotypes calculated separately among populations within each region

Table 3.3: Numbers of eurytopic haplotypes found within individual sampling locations

Haplotype **	Location *																			
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
	BAMA	*BOPI*	*BOSO*	*EEZO*	*GEKA*	*IARA*	*IIAA*	*IIAE*	*IIAW*	*IILA*	*KEDN*	*KEKA*	*KEKU*	*KENO*	*KUSL*	*NESN*	*RIKA*	*SIMU*	*SINI*	*TOBA*
1	1	1	4																	
2	1											1								
3	1					3							2						1	
4	8		1	2											2					
5		3											1				1			
6		3	1			1														
7		1	4	2																
8			1			1														
9				2		3						2								
10				1					3				1					7	3	
11				2	1												2			
12				1													1			
13				1			1													
14				1										1						
15				1						2					1		1			
16				1							1	2	1	5						
17				1					1		1		1	2			1			
18											6	2			2					
19												3	1							
20												4		2			1			
21													6						2	
22													1						1	
23																	1			1
24																	1	3		

* For full description of sampling locations, see Table 2.1 and associated text.

** Haplotypes are arbitrarily numbered.

ranged from 20% to 100% for the sample of eastern forests (*i.e.* when multiple locations within forests were collapsed) and from 86% to 100% for the Zaïrian sample.

These differences are supported statistically. Independent t-tests showed that endemism among Zaïrian forests was significantly greater than endemism within eastern forests (17 degrees of freedom; $p = 0.003$). Within the five Zaïrian locations, the number of endemic haplotypes was greater than the number of eurytopic haplotypes (paired t-test with 4 degrees of freedom; $p = 0.004$). Within the 14 eastern populations, however, the number of endemic haplotypes was not statistically different from the number of eurytopic haplotypes (paired t-test with 13 degrees of freedom; $p = 0.87$). Examination of the data reveals that the number of endemic haplotypes is consistently low for locations within Kibale forest. This is probably an artifact of uneven geographic sampling, since Kibale was sampled on a finer geographic scale than other forests; Kibale haplotypes were defined as eurytopic even if they appeared only in another sampling location within Kibale. To correct this bias, the Kibale sample was collapsed for the purposes of analysis into a single "Kibale Forest" population. The number of endemic haplotypes within eastern forests was still not significantly greater than the number of eurytopic haplotypes (paired t-test with 10 degrees of freedom; $p = 0.72$), and endemism within Zaïre was still higher than endemism within eastern forests (independent t-test with 14 degrees of freedom; $p = 0.004$).

The apportionment of eastern chimpanzee diversity

The framework for the following analysis is described in Excoffier, Smouse and Quattro (1992), and will be referred to as AMOVA, or analysis of molecular variance, following the terminology of the authors. AMOVA is built on a population genetic framework first described by Wright (Wright, 1951; Wright, 1965; Wright, 1969) involving the calculation of *F*-statistics, or fixation indices, which measure the deviation of genotype frequencies in subdivided populations from expectation under panmixia. The

approach is adaptable to a range of genetic systems (Cockerham, 1969; Nei, 1977; Nei, 1987), and is usually structured as a three-tiered hierarchy of genetic diversity components (Lewontin, 1972; Weir and Cockerham, 1984).

AMOVA partitions total genetic variation into variation a) within populations, b) among populations within regions, and c) among regions, where regions are exclusive groups of populations, defined by *a priori* criteria. The proportion of total genetic variation attributable to each of these levels indicates the nature and extent of subdivisioning present. For example, Lewontin's (1972) analysis of human genetic variation at the protein level attributed only 6% of the total variation within the species to variation between regions (races), thereby suggesting that genetic differences among human races are minor. Similar analyses of mitochondrial DNA have supported these findings by indicating that the largest component of human genetic variation lies within populations (Stoneking *et. al.*, 1990; Excoffier, Smouse and Quattro, 1992; Barbujani *et. al.*, 1995).

Excoffier, Smouse and Quattro (1992) adapt this approach to incorporate information about molecular distances among haplotypes. This information is valuable in that it can reveal "hidden" genetic variation, ignored by traditional *F*-statistic approaches which assume equidistance among alleles. To signify this departure from previous approaches, Excoffier, Smouse and Quattro (1992) define "ϕ-statistics." The authors follow Cockerham (1969) and propose the statistics ϕ_{st}, ϕ_{ct} and ϕ_{sc} as molecular analogs of fixation indices within populations, among populations within regions, and among regions, respectively.

This approach was applied to the sample of eastern chimpanzees using the computer program AMOVA (Excoffier, Smouse and Quattro, 1992). Populations were defined in two ways: 1) as the 19 sampling locations for which the geographic provenance of the animals was certain and 2) as "forests" *(i.e.* separate locations within Kibale, Semliki and Budongo were collapsed). The former definition of populations is only as arbitrary as the

geographic sampling strategy used in the study; it was chosen to minimize further *a priori* assumptions about what constitutes a population. The latter definition reflects the notion that chimpanzees most likely migrate freely within forests. This definition, however, assumes a population substructure (or lack of one) and may bias the results in favor of high intra-population variation. Regions were defined as Zaïrian forests and Eastern forests. This definition is relatively arbitrary. The effects of defining populations and regions in alternate ways are explored below.

Two matrices of inter-allelic distance were analyzed. The first corresponds to Excoffier, Smouse and Quattro's (1992) "D_2" matrix in which all haplotypes are considered equidistant (all distinct haplotypes differ by an equal genetic distance, regardless of varying amounts of nucleotide sequence difference). The resulting "multiallelic" ϕ-statistics reduce to the traditional F-statistics, making the corresponding variance components directly comparable with previous studies which have adopted a multiallelic approach (Lewontin, 1972; Long, 1986; Stoneking *et. al.*, 1990; Barbujani *et. al.*, 1995). The second matrix describes euclidean distances among haplotypes, inferred from the number of nucleotide differences between pairs of sequences without regard to the nature of these differences (e.g. transitions or transversions). This phenetic matrix is equivalent to Excoffier, Smouse and Quattro's "D_1" (haplotypic) matrix, which, in the human case, yields results virtually identical to those derived from cladistic distances among haplotypes (Excoffier, Smouse and Quattro, 1992).

Results are presented in Table 3.4, along with results from Excoffier, Smouse and Quattro's analysis of a world sample of 672 human mitochondrial RFLP's. Despite the fact that the geographic ranges of eastern chimpanzee populations are small compared to those of the human populations defined by Excoffier, Smouse and Quattro, humans and eastern chimpanzees display similar proportions of within-population variance (approximately 80-90%). However, the proportion of variance accounted for by variance

Table 3.4: Hierarchical analysis of molecular variance for humans and East African chimpanzees

*Homo sapiens**

		Multiallelic variance†				Haplotypic variance‡			
		Observed partition				Observed partition			
Variance component		Variance	% total	p^a	ϕ-statistics	Variance	% total	p^a	ϕ-statistics
Among regions	σ_a^2	0.055	15.73	0.0080	$\phi_{CT} = 0.157$	0.134	21.12	0.0020	$\phi_{CT} = 0.211$
Among populations/regions	σ_b^2	0.013	3.59	<0.0001	$\phi_{SC} = 0.043$	0.022	3.49	<0.0001	$\phi_{SC} = 0.022$
Within populations	σ_c^2	0.281	80.68	<0.0001	$\phi_{ST} = 0.193$	0.478	74.39	<0.0001	$\phi_{ST} = 0.478$

*Pan troglodytes schweinfurthii (populations defined as sampling locations)***

		Multiallelic variance†				Haplotypic variance‡			
		Observed partition				Observed partition			
Variance component		Variance	% total	p^a	ϕ-statistics	Variance	% total	p^a	ϕ-statistics
Among regions	σ_a^2	0.001	0.11	0.3826	$\phi_{CT} = 0.001$	0.081	2.01	0.0859	$\phi_{CT} = 0.020$
Among populations/regions	σ_b^2	0.064	12.80	<0.0010	$\phi_{SC} = 0.128$	0.725	18.01	<0.0010	$\phi_{SC} = 0.184$
Within populations	σ_c^2	0.432	87.09	<0.0010	$\phi_{ST} = 0.129$	3.219	79.98	<0.0010	$\phi_{ST} = 0.200$

* World sample of 672 human mtDNAs (restriction digests); data taken from Excoffier, Smouse and Quattro (1992)

** 255 eastern chimpanzee mitochondrial control region sequences from 19 forest locations

† Multiallelic variances were calculated assuming all haplotypes were equidistant.

‡ Haplotypic variances were calculated using a matrix of distances among haplotypes, based on DNA sequences (uncorrected for multiple substitutions).

[a] Probability of having a more extreme variance component and phi-statistic than the observed values by chance alone. Probabilities calculated by a random permutation procedure (see text and Excoffier, Smouse and Quattro, 1992).

Table 3.4 (continued): Hierarchical analysis of molecular variance for humans and East African chimpanzees

*Pan troglodytes schweinfurthii (populations defined as forests)****

		Multiallelic variance†				Haplotypic variance‡			
		Observed partition				Observed partition			
Variance component		Variance	% total	p^a	ϕ-statistics	Variance	% total	p^a	ϕ-statistics
Among regions	σ_a^2	0.000	0.00	0.8794	$\phi_{CT} = 0.000$	0.076	1.88	0.1308	$\phi_{CT} = 0.019$
Among populations/regions	σ_b^2	0.051	10.23	<0.0010	$\phi_{SC} = 0.102$	0.561	13.87	<0.0010	$\phi_{SC} = 0.141$
Within populations	σ_c^2	0.447	89.77	<0.0010	$\phi_{ST} = 0.102$	3.409	84.25	<0.0010	$\phi_{ST} = 0.158$

* World sample of 672 human mtDNAs (restriction digests); data taken from Excoffier, Smouse and Quattro (1992)

** 255 eastern chimpanzee mitochondrial control region sequences from 19 forest locations

† Multiallelic variances were calculated assuming all haplotypes were equidistant.

‡ Haplotypic variances were calculated using a matrix of distances among haplotypes, based on DNA sequences (uncorrected for multiple substitutions).

Probability of having a more extreme variance component and phi-statistic than the observed values by chance alone. Probabilities calculated by a random permutation procedure (see text and Excoffier, Smouse and Quattro, 1992).

*** Sampling locations within forests were collapsed.

chimpanzees as within humans (12-18% in chimpanzees, versus 3.5-3.6% in humans). Genetic differences between chimpanzee populations are therefore relatively larger than those between human populations, although chimpanzee populations are separated by much smaller distances. In contrast, inter-regional variation in eastern chimpanzees is relatively low, accounting for a maximum of only two percent of the variation within the subspecies. Regional differences in humans are almost ten times as great (16-21%), as might be expected from the fact that human regions, as defined by Excoffier, Smouse and Quattro (1992), correspond to geographic divisions on a continental scale. Chimpanzee regions, defined here as "Zaïrian forests" and "eastern forests," are apparently not equivalent to human regions, and do not appear to have any equivalent biological meaning for eastern chimpanzees. Nevertheless, the Excoffier, Smouse and Quattro analysis of human RFLP data should be viewed cautiously. The degree of regional or "racial" difference documented by by Excoffier, Smouse and Quattro (roughly 20%) is more than twice as great as that found by Cann *et. al.* (1987), also using human mitochondrial RFLP data (8% at most; see Cann *et. al.*, 1987, Table 1). The reasons for this discrepancy are not clear.

Null distributions of variances were generated by random permutation of the data as a test of the hypothesis that the observed variance components resulted from chance (sampling error), rather than from real population subdivisioning (Excoffier, Smouse and Quattro, 1992). This approach is powerful in that it is non-parametric and thus makes minimal *a priori* distributional assumptions about the data. One thousand random permutations were performed on a) entire populations across regions, keeping numbers of populations within each region constant, b) individuals within regions, keeping sample sizes of individuals within each region constant, and c) individuals within populations, keeping sample sizes of individuals within each population constant. Observed variances were compared to the simulated null distributions to assess the probability of obtaining a

variance as extreme as the empirical variance by chance alone.

Representative results (from the "sampling location" definition of populations) are shown in Figure 3.1. Results from the "forest" definition of populations did not differ from those shown. Open bars represent the null distribution generated by simulation, and closed markers represent the location of the empirical value of each variance component. Variance among regions is not significantly different from chance, either in the multiallelic or haplotypic case. Incorporation of information about molecular distances among haplotypes augments regional differentiation, but not to the level of statistical significance. This result contrasts with the human mitochondrial data, which indicate significant inter-regional variation accounting for as much as 21% of the variation within humans (Excoffier, Smouse and Quattro, 1992), although less than half that amount according to Cann *et. al.* (1987). Greater inter-regional variation in humans, if real, might reflect the unusually large range of the human species. Humans can be examined on a vastly greater geographic scale than can eastern chimpanzees. Eastern chimpanzee regions, as defined here, are fundamentally different from human regions, which were defined as continents.

Eastern chimpanzees show a significant excess of among-population, within region variance over what would be expected by chance alone. Genetic differentiation among populations is therefore greater than that attributable to sampling error. As previously stated, the degree of population differentiation in eastern chimpanzees is greater than that in humans despite the fact that human populations, as defined by Excoffier, Smouse and Quattro (1992), are separated by much larger geographic distances. This observation again supports the notion that humans are extraordinarily vagile. Finally, eastern chimpanzees show a statistically significant lack of within-population genetic diversity over what would be expected by chance alone. Like humans, individual eastern chimpanzee populations as defined in the present study contain most, but not all, of the genetic variation within the taxon.

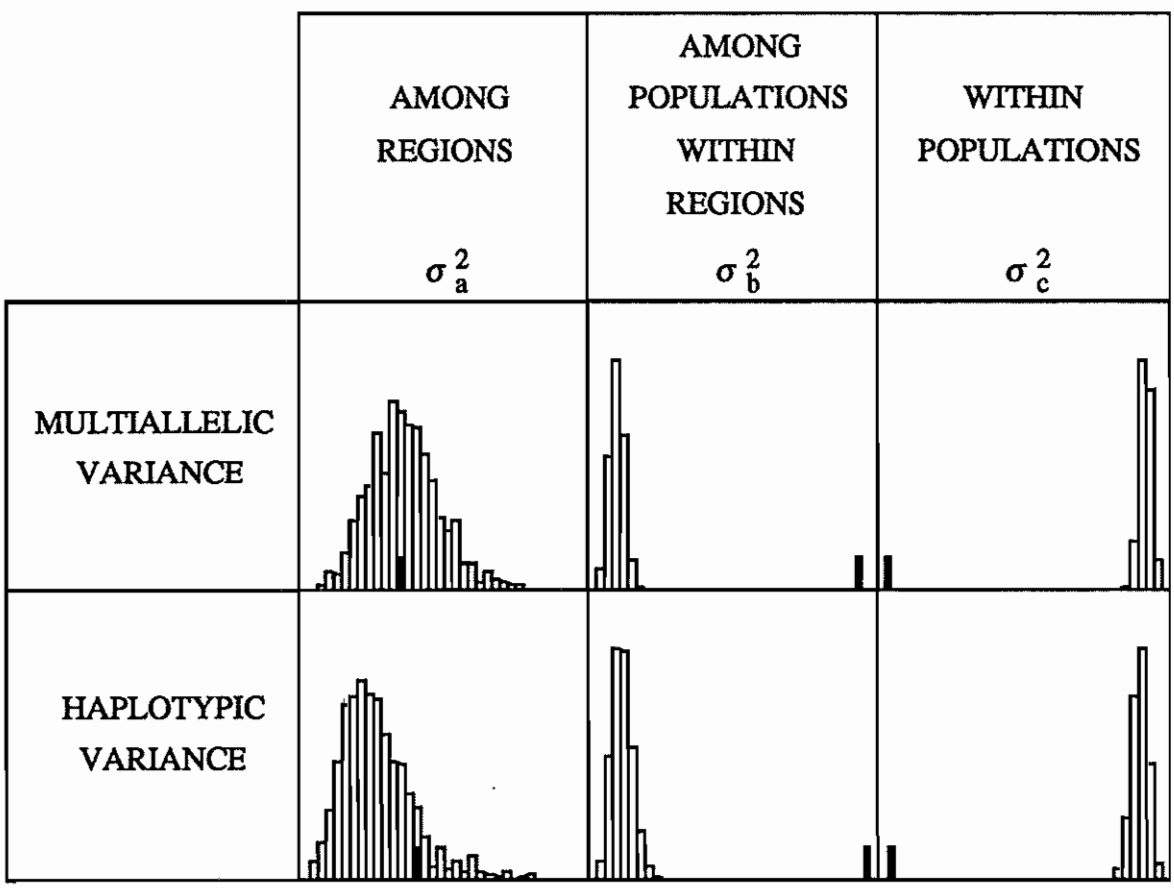


Figure 3.1: Null distributions of multiallelic and haplotypic variance components at three hierarchical levels. Open bars show the distribution of variances generated by 1000 random permutations of a) populations across regions (keeping numbers of populations within each region constant), b) individuals within regions (keeping numbers of individuals within each region constant), and c) individuals within populations (keeping numbers of individuals within each population constant). Solid markers show the positions of observed empirical values. Multiallelic variance components were generated assuming equidistance among alleles; haplotypic variance components were generated using phenetic distances among haplotypes. Two regions were considered: Eastern forests, and Zaïrian forests. Distributions were generated using the program AMOVA (Excoffier et. al., 1992).

In general, the degree of population subdivisioning evident in eastern chimpanzees is small. The multiallelic value of $\phi_{ST} = 0.129$ (equivalent to Wright's F_{ST}) is low in comparison to taxa which show strong subdivisioning (Allendorf, 1983; Avise, 1994). F_{ST} bears a direct relationship to *Nm*, the product of a population's effective size and its average per generation migration rate (Wright, 1969; Takahata and Palumbi, 1985). The product *Nm* thus estimates the absolute number of individuals which migrate among populations per generation. Although *N* and *m* are difficult to measure separately, their product is easily obtained from mitochondrial data as a single value: $Nm = (1 - F_{ST})/2F_{ST}$ (Takahata and Palumbi, 1985). Takahata and Palumbi's (1985) modified formula for the calculation of *Nm* from extranuclear genetic data using $F_{ST} = 0.129$ (Table 3.4) yields an estimate of $Nm = 3.4$ migrants exchanged, on average, between populations per generation in eastern chimpanzees. This degree of migration is high; a value of *Nm* of approximately 1 is theoretically sufficient to prevent population subdivisioning due to drift alone (Allendorf, 1983; Avise, 1994). Accurate estimates of *Nm* can be obtained only when many genetic systems are examined (Allendorf, 1983). Nevertheless, the present data strongly suggest that migration among eastern chimpanzee populations, as defined for the study, has been extensive.

The results described above did not differ when alternate definitions of populations and regions were tested. Results for the "forest" definition of populations are presented in Table 3.4. The proportions of among region, among population/within region, and within region variance were, respectively, 0.00 %, 10.23% and 89.77% (multiallelic) and 1.88%, 13.87% and 84.25% (haplotypic). The multiallelic value of ϕ_{ST} by this analysis was 0.102, implying an average per generation migration rate of 4.4 individuals. Regions were alternatively defined as refuge populations (Rwenzori and Semliki forests plus all Zaïrian populations) and non-refuge populations (all other forests). This definition actually decreased the proportion of inter-regional variation in both the multiallelic (0.00%) and haplotypic (1.44%) cases. Other regional definitions explored were "northern forests"

(forests North of the Kasinga channel in Uganda) and "southern forests" (forests south of the Kasinga channel in Uganda), and a regional definition in which all Zaïrian populations separated by large rivers were defined as separate regions. The proportion of inter-regional variation was no different than that described for the original definition of regions (east versus Zaïre) in either case.

In general, therefore, the proportion of variance among regions never was greater than approximately 2%, despite testing of a variety of alternate definitions. Similarly, among-population variance ranged between approximately 10% and 19% by any alternative definition of populations. ϕ_{ST} was never greater than approximately 0.2. These results suggest that imposing population structure on the data *a priori* will not alter the conclusion that the eastern chimpanzee subspecies shows only moderate population subdivisioning and has experienced extensive migration.

Tajima's neutrality test

Given that a significant proportion of genetic variation within east African chimpanzees is attributable to variation among populations, it is justified to test hypotheses pertaining to the causes of that variation. Such hypotheses, described below and in Chapter 1, rely on the notion that microevolutionary genetic differentiation among populations is mediated by neutral processes--specifically the fixation of point mutations or entire alleles through genetic drift. It is therefore important to test the assumption that the data do, in fact, conform to a model of neutral mutation.

Tajima (1989) describes a statistical test of neutral mutation, which is derived from an infinite sites model of molecular evolution (Kimura, 1969). Tajima's test compares the number of segregating (polymorphic) nucleotide sites in a sample to the average number of nucleotide differences in that sample. Under selective neutrality, these two measures are theoretically correlated to a degree which increases with sample size (of DNA sequences).

Significant departures from expected values of Tajima's test statistic, D, would indicate rejection of the null model of selective neutrality, and might point to selection or other disequilibrium-producing forces, such as recent demographic change (Tajima, 1989).

Results from Tajima's test are presented in Table 3.5. A significant departure from neutral expectation was detected for only one population (SIMU). Values of D for other populations were well within the bounds of statistical expectation, as were values for eastern forests, Zaïrian forests, and the total population combined. This observation suggests that the SIMU result likely represents type-1 error, and that a model of selective neutrality cannot be rejected for the data as a whole. Eastern chimpanzees differ in this respect from humans, for which significant departures from neutrality in mitochondrial DNA have been detected using Tajima's test (Merriwether *et. al.*, 1991; Barbujani *et. al.*, 1995). The lack of such a trend in chimpanzees is surprising given the analysis in Chapter 4 supporting a hypothesis of recent demographic expansion in the subspecies, which can lead to deviations of the D statistic from neutral expectation. Nevertheless, results from Tajima's test validate the use of models of interpopulational genetic differentiation which assume the predominance of selectively-neutral evolutionary forces.

Biogeographic predictors of diversity and endemism

Areas of Eastern forests were obtained from Howard (1991) and Weber (1987), and are listed in Table 2.1. The area of Zaïrian forest suitable for eastern chimpanzees has been estimated by Teleki (1989) as 470,000 km^2. For the purposes of this analysis, the area of all Zaïrian forest locations was defined as equivalent to the area of the largest eastern forest (NESN; 1140 km^2) to preserve approximate normality in the distribution of area measurements. For each sampling location, a distance was calculated to the "source" population. On the basis of biogeographic evidence, source populations (Pleistocene refugia) sampled in this study are assumed to be the five Zaïrian populations (Grubb, 1982;

Table 3.5: Tajima's neutrality test applied to individual sampling locations

	Location*	n	S**	$\hat{k}$†	D‡	p[a]
1	BAMA	13	14	2.897	-1.617	ns
2	BOPI	12	21	5.424	-1.084	ns
3	BOSO	15	14	5.333	1.009	ns
4	EEZO	26	48	7.203	-1.685	ns
5	GEKA	19	22	7.041	0.482	ns
6	IARA	12	20	5.227	-1.033	ns
7	IIAA	13	33	10.077	-0.256	ns
8	IIAE	12	14	4.091	-0.555	ns
9	IIAW	13	21	7.641	0.609	ns
10	IILA	17	22	6.485	-0.014	ns
11	KEDN	14	19	6.747	0.583	ns
12	KEKA	15	10	3.333	0.338	ns
13	KEKU	14	23	7.571	0.215	ns
14	KENO	13	20	7.000	0.404	ns
15	KUSL	13	21	5.385	-0.964	ns
16	NESN	13	19	6.359	0.180	ns
17	RIKA	13	29	9.167	-0.092	ns
18	SIMU	13	14	6.744	2.236	< 0.05
19	SINI	11	22	7.600	0.062	ns
20	TOBA	10	19	6.089	-0.516	ns
Eastern Forests		171	68	7.478	-1.146	ns
Zaïrian Forests		65	58	7.817	-1.226	ns
Total		281	90	7.637	-1.422	ns

* For full description of sampling locations, see Table 2.1 and associated text.

** Number of segregating sites

† average number of pairwise nucleotide differences

‡ Tajima's (1989a) test statistic

[a] Probabilities calculated from confidence limits of D given in Tajima (1983)

Colyn, 1991; Colyn, Gautier-Hion and Verheyen, 1991) and Rwenzori (Hamilton, 1976; Hamilton, 1981; Struhsaker, 1981; Rodgers, Owen and Homewood, 1982). Distances from these forests to the source was therefore defined as zero. Distances from other populations to the source were measured as great-circle distances to the Rwenzori population (RIKA), based on an estimated diameter of the Earth of 12,756 km (Skinner and Parker, 1987). RIKA was chosen because it is the closest "source" population to all non-refuge forests (Struhsaker, 1981; Rodgers, Owen and Homewood, 1982). Populations also varied with respect to their geographic centrality within the study area. Isolation was therefore quantified as the mean great-circle distance of a population to all other populations in the study. Geographic measurements used in the following analyses are presented in Table 3.6.

Figure 3.2 shows the relationship of Shannon-Wiener diversity and endemism to forest area, distance to the source, and isolation. Diversity shows the predicted positive relationship to each of the geographic variables. However, in no case is this relationship statistically significant (area: $r^2 = 0.04$, $p = 0.43$; distance: $r^2 = 0.07$, $p = 0.29$; isolation: $r^2 = 0.08$, $p = 0.23$). The relationship between endemism and area is similarly positive, but not significant ($r^2 = 0.21$, $p = 0.07$). Contrary to the expectations of Pleistocene refuge theory, endemism correlates positively with both distance to the source and isolation. Although the relationship between endemism and distance is not significant ($r^2 = 0.03$, $p = 0.52$), the relationship between endemism and isolation is significant ($r^2 = 0.34$, $p = 0.02$), although r^2 is unimpressive.

Figure 3.3 shows the relationship of the four remaining genetic diversity measures to these same three biogeographic variables. The relationship between area and all four measures is in the predicted positive direction, but is not significant in any case (gene diversity: $r^2 = 0.05$, $p = 0.35$; nucleotide diversity: $r^2 = 0.05$, $p = 0.38$; modal sequence difference: $r^2 = 0.07$, $p = 0.27$; maximum sequence difference: $r^2 = 0.00$, $p = 0.86$). No

Table 3.6: Geographic measurements used in the analysis of genetic diversity

	*Location**	*Area (km^2)***	*Distance to the "source" (km)****	*Isolation (km)†*
1	BAMA	365	130	233
2	BOPI	793	238	319
3	BOSO	793	209	293
4	EEZO			
5	GEKA	32	583	617
6	IARA	87	45	184
7	IIAA	470,000	208	257
8	IIAE	470,000	295	312
9	IIAW	470,000	297	313
10	IILA	470,000	186	242
11	KEDN	560	44	189
12	KEKA	560	28	178
13	KEKU	560	26	179
14	KENO	560	36	181
15	KUSL	580	111	227
16	NESN	1140	335	389
17	RIKA	996	0	176
18	SIMU	212	23	177
19	SINI	212	31	178
20	TOBA	470,000	339	353

* See Figure 2.1 for map of sampling locations.

** In the analyses, areas of Zaïrian forests were defined as the area of the largest insular forest (NESN) to preserve approximate normality in the distribution of area measurements.

*** RIKA was defined as the "source" because it is the refuge location closest to all eastern insular forests.

† Isolation was measured as the mean distance of a population from all other populations in the study.

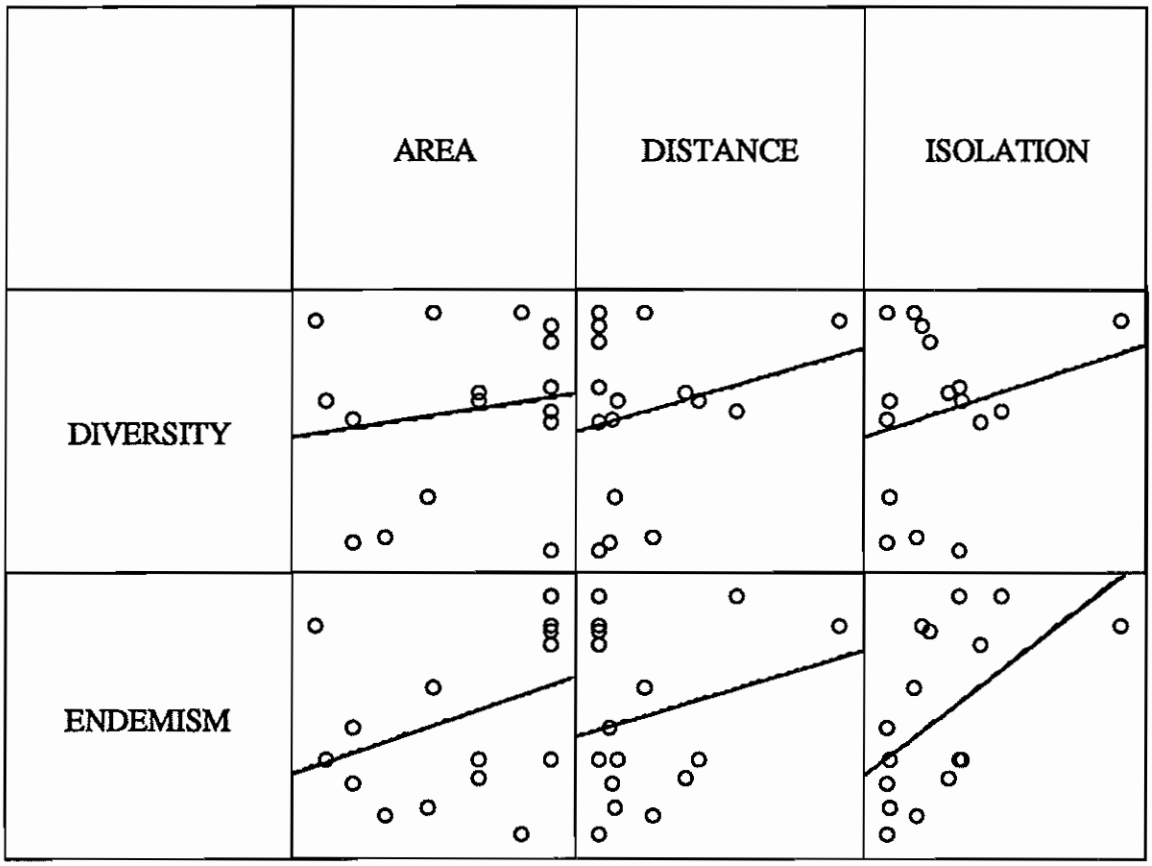


Figure 3.2: Biogeography theory and chimpanzee genetics. The figure shows the relationship between two critical parameters of biogeography theory and three potential predictors of these parameters for 19 chimpanzee populations. The biogeographic parameters are diversity (Shannon-Weiner Index of haplotype diversity) and endemism (percentage of endemic haplotypes). The geographic predictors are forest area, distance from the "source" and relative isolation. Areas of Zaïrian sampling locations were defined as equivalent to the maximum area of any eastern forest (1140 km^2). For measures of endemism, separate sampling locations within Kibale forest were collapsed because of bias introduced by concentrated geographic sampling (see text). Axes are linear: Diversity: 1.2-2.4; Endemism: 0-100%; Area: 32-1150 km^2; Distance: 0- 600 km; Isolation: 170-620 km. Geographic parameters used in the regressions are given in Table 3.6.

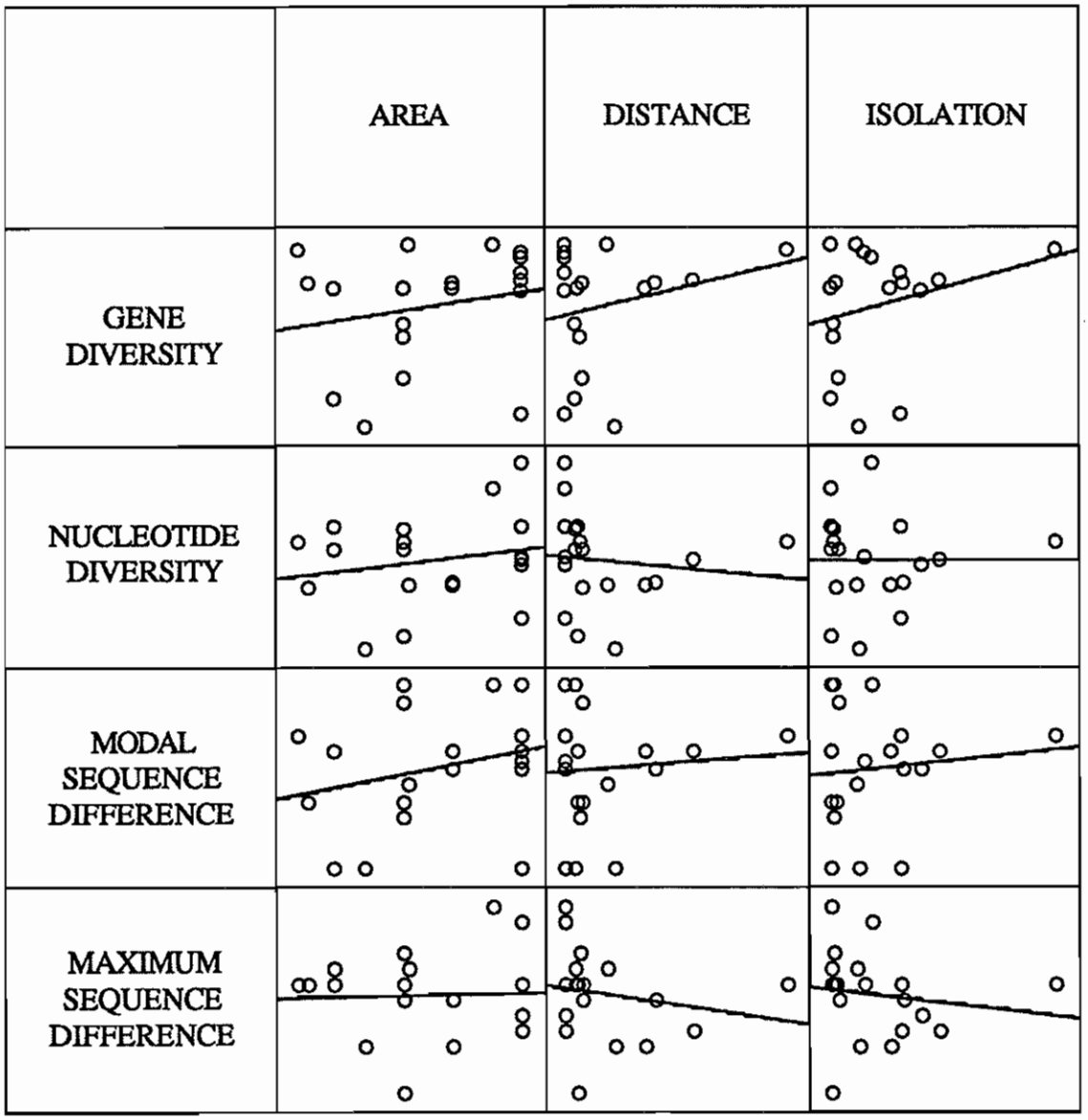


Figure 3.3: Geography and genetic diversity. The figure shows the relationship between three geographic parameters and four measures of genetic diversity for 19 chimpanzee populations. The geographic parameters are forest area, distance from the "source," and relative geographic isolation. The genetic diversity measures are Nei's (1973) gene diversity, Nei and Li's (1979) nucleotide diversity, and modal and maximum pairwise distances among sequences within each population. Axes are linear: Gene diversity: 0-1; Nucleotide diversity: 0-3%; Modal sequence difference: 0-3%; Maximum sequence difference: 2-5%; Area: 0-1150 km^2; Distance: 0-600 km; Isolation: 170-620 km. Geographic parameters used in the regressions are given in Table 3.6.

consistent trend is observed between any genetic diversity measure and distance (gene diversity: $r^2 = 0.05$, $p = 0.36$; nucleotide diversity: $r^2 = 0.00$, $p = 0.90$; modal sequence difference: $r^2 = 0.00$, $p = 0.81$; maximum sequence difference: $r^2 = 0.03$, $p = 0.45$), or between any diversity measure and isolation (gene diversity: $r^2 = 0.06$, $p = 0.33$; nucleotide diversity: $r^2 = 0.00$, $p = 0.94$; modal sequence difference: $r^2 = 0.01$, $p = 0.71$; maximum sequence difference: $r^2 = 0.03$, $p = 0.50$). Multivariate examination of each of the relationships described above (*i.e.* treating area, distance and isolation as interrelated variables in a multiple regression) did not change the results.

Sampling error could account for the lack of the insular biogeographic model's predictive power. In addition, the lack of a significant association between area and diversity/endemism may reflect error in the measurement of forest area, or in the use of present-day forest areas as surrogates for "actual" forest areas prior to recent deforestation (Hamilton, 1984; Hamilton, Taylor and Vogel, 1986). The possibility also exists that the predictive biogeographic model analyzed was inaccurate. Chimpanzees may have experienced a biogeographic history different from that suggested by refuge models formulated from data on the distributions of other taxa (Diamond, 1979; Kingdon, 1981; Struhsaker, 1981; Rodgers, Owen and Homewood, 1982; Colyn, 1991; Colyn, Gautier-Hion and Verheyen, 1991). If so, then distance from the "real" chimpanzee refuge might still be a valid predictor of genetic diversity and endemism. Biogeographic hypotheses about the locations of Pleistocene refugia are fundamentally *a posteriori*, in that they often define refugia specifically as those areas in which diversity and endemism are greatest (Endler, 1982; Mayr and O'Hara, 1986). This study therefore asks whether any single location within the bounds of the study area has properties which conform to the expectations of a Pleistocene refuge. Such an area would presumably be maximally genetically diverse, and would contain a large proportion of endemic haplotypes. Both diversity and endemism would decline with distance from this area in all directions.

To locate such an area in space, a rectangular grid of geographic coordinates was

constructed between longitudes 26.0° and 32.0°, and between latitudes -5.0° and 3.0 °. Points were regularly spaced every 0.2° (longitudinal and latitudinal) within these boundaries. Great-circle distances were measured between each of the resulting 1066 locations and the 19 forest locations from which chimpanzee DNA sequences were obtained. The Pearson's correlation coefficient was then calculated for the association between endemism/diversity and distance of each forest location from that point. The magnitude of the resulting correlation coefficient indicates the strength of the association; its sign indicates the direction of the association. By the predefined criteria of this analysis, "refugia" should correspond to locations from which the correlation is maximally negative.

Results are presented in Figures 3.4 and 3.5 as a series of three-dimensional surfaces, with z-axes representing the value of the Pearson correlation coefficient calculated as described above from each point defined by an x (longitudinal) and y (latitudinal) coordinate. Figure 3.4 shows a clear southwest-northeast cline in the strength of association between Shannon-Wiener diversity and distance, with a maximally-negative r (r_{-max}) of -0.26 at longitude 29.0° and latitude -5.0°. The strength of association between endemism and distance follows a very similar cline, with r_{-max} = -0.61 at longitude 26.0° and latitude -2.8°. Figure 3.5 presents results of the same analysis for the association between distance and gene diversity (r_{-max} = -0.26; longitude 27.8°, latitude -5.0°), nucleotide diversity (r_{-max} = -0.27; longitude 27.4°, latitude -1.6°), modal sequence difference (r_{-max} = -0.17; longitude 28.0°, latitude -4.8°), and maximum sequence difference (r_{-max} = -0.27; longitude 28.8°, latitude -0.2°).

The latitudinal and longitudinal concordance of r_{-max} for all measures is intriguing, as is its general decrease in strength and negativity with distance from points in the southwest of the study area. The geometric center of mass of r_{-max} for all five diversity measures and endemism lies at longitude 27.8°, latitude -3.2°, which is located in Zaïre, approximately 100 km due west of the northern shore of Lake Tanganiyka. These results

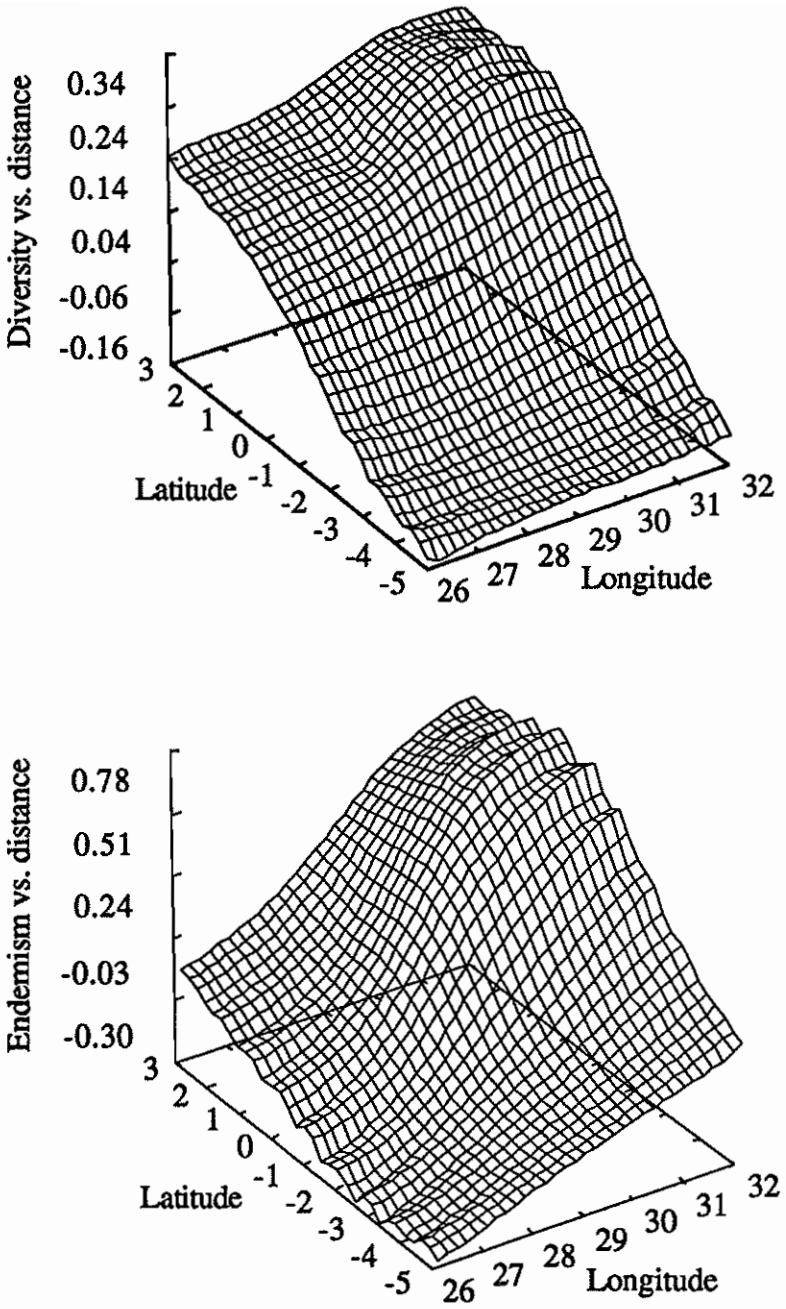


Figure 3.4: Topographic representation of the strength of association between distance and diversity (Shannon-Wiener Index; upper panel) and between distance and endemism (percent endemic haplotypes; lower panel) as a function of location. For each point on a 0.2° latitudinal by 0.2° longitudinal grid within the study area, a Pearson's correlation coefficient was calculated for the relationship between diversity/endemism and distance from that point for 19 forest locations. The Z axis displays the resulting correlation coefficients, unsquared to preserve negativity.

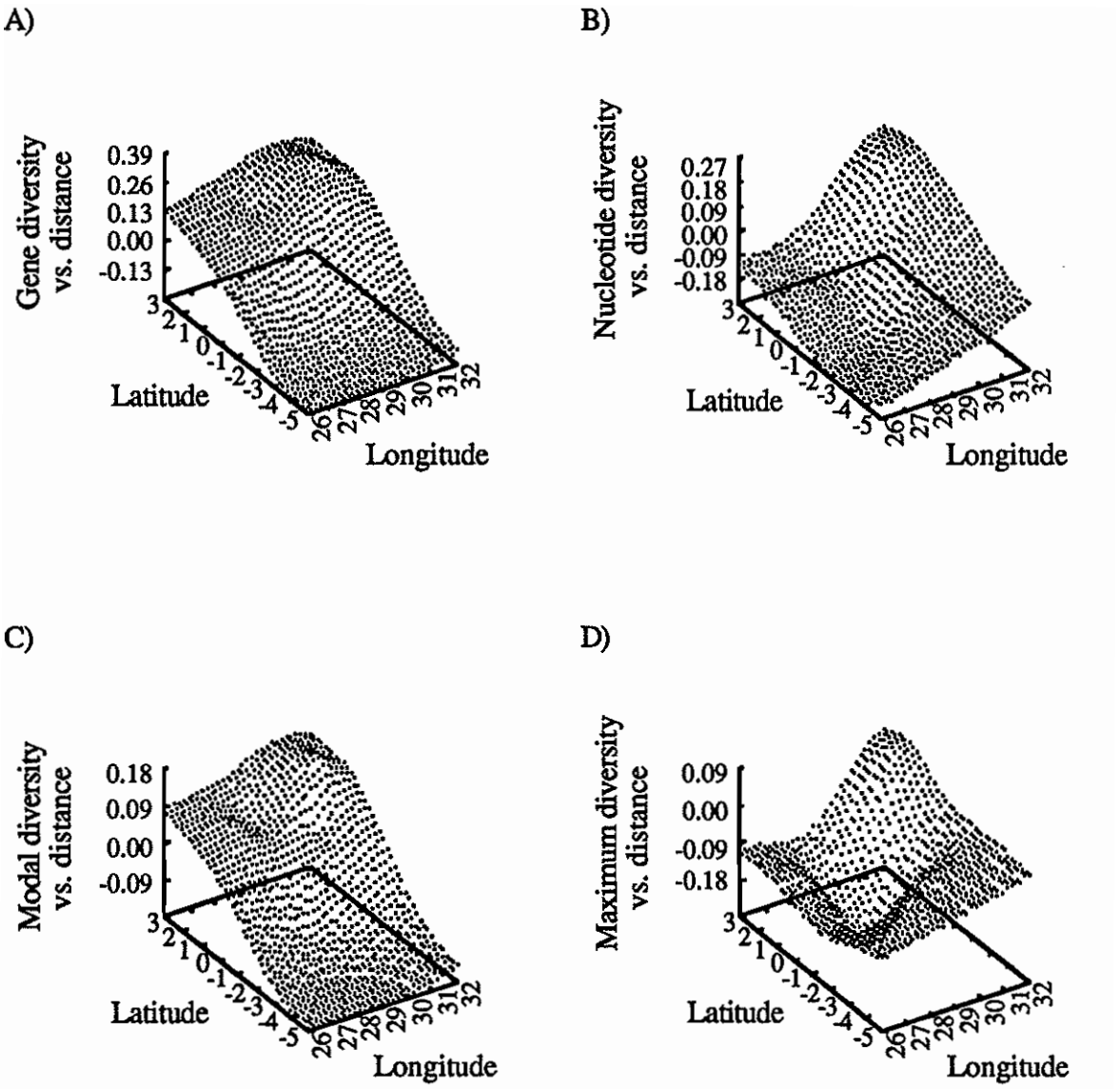


Figure 3.5: Topographic representations of the strength of association between distance and four measures of genetic diversity as a function of location. Diversity measures are A) Nei's (1987) gene diversity, B) Nei's (1987) nucleotide diversity, C) the modal pairwise sequence difference between haplotypes within each population and D) the maximum pairwise sequence difference between haplotypes within each population. Z axes are Pearson's correlation coefficients, unsquared to preserve negativity.

are, however, entirely dependent on the geometric arrangement of sampling locations. For example, a southwest-northeast cline of correlational strength would never be observed for forests situated on a perpendicular northwest-southeast archipelago. The precise location of $r_{\text{-max}}$ would be unbiased only if sampling locations were evenly-distributed across the landscape. More importantly, no value of $r_{\text{-max}}$ was notably high. Distance from any location explains a maximum of only 37% of the variation in any diversity measure ($r_{\text{-max}}^2$ = 0.372 for the relationship between distance and endemism). Distance from any single "source" population is therefore likely to be a poor predictor of genetic diversity, regardless of the biogeographic model tested.

Biogeographic predictors of genetic distance

Figure 3.6 presents a map depicting the extent of haplotype sharing among sampling locations. Linkages are drawn as straight lines, and are proportional in width to the numbers of haplotypes shared between locations. Populations may share haplotypes either because of common ancestry or recent gene flow. This analysis assumes an "infinite sites" model of molecular evolution and therefore disregards evolutionary convergence. Linkages between populations based on haplotype sharing therefore provide a time-averaged measure of genetic interconnectedness, but do not give an indication of the extent of gene flow occurring today.

Several properties of this map are immediately suggestive. First, haplotypes are shared even among populations which are geographically very distant. The maximal distance over which any haplotype is shared is 583 km, between GEKA and RIKA populations. This distance is, however, less than the maximum geographic distance among populations (772 km), suggesting that gene flow may be limited over very large distances. Second, haplotype sharing appears greatest among populations separated by small distances. Third, haplotypes are shared between Zaïrian and eastern locations. Finally,

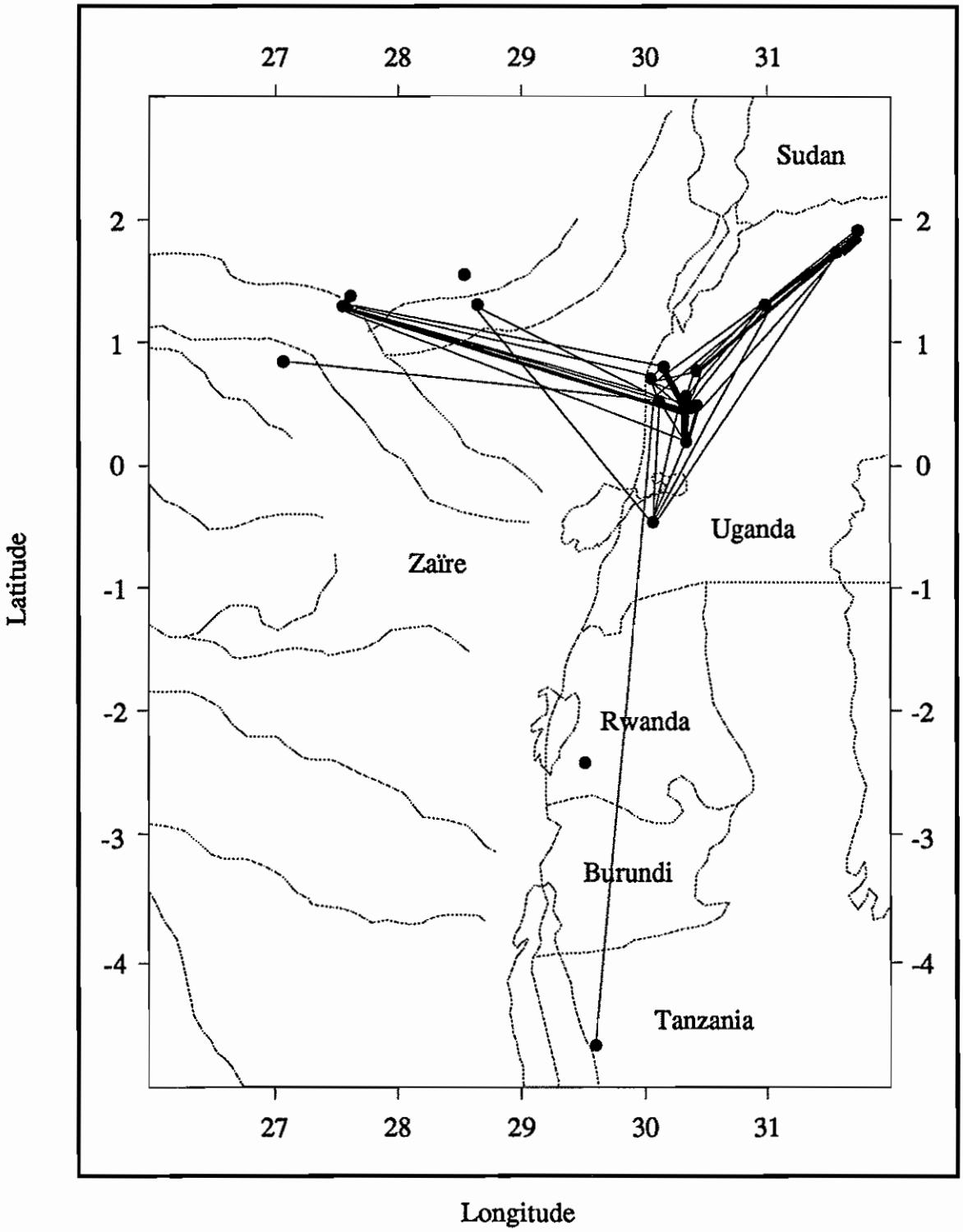


Figure 3.6: Genetic linkages between sampling locations, based on 24 eurytopic haplotypes. Individual sampling locations (n=19) are represented by closed circles. Lines represent the sharing of a haplotype between locations. Line widths are proportional to numbers of haplotypes shared.

haplotype sharing among eastern forests is extensive, while haplotype sharing among Zaïrian populations is nonexistent. Pairs of Zaïrian populations on opposite sides of rivers (IIAA and IILA; IIAE and IIAW) share no haplotypes, even though some are separated by distances less than the smallest distances between eastern populations.

Linkages based on eurytopic haplotypes are useful visually, but not statistically. The number of haplotypes shared between locations is a direct function of sample size. Since only 24 eurytopic haplotypes were identified, strong inferences should not be made from the linkage pattern shown in Figure 3.6. The absence of any shared haplotypes within Zaïre, for example, likely reflects a relatively high degree of population subdivisioning within Zaïre, rather than complete genetic isolation. This conclusion is strengthened by the fact that haplotypes are shared between Zaïrian and eastern populations. With more extensive genetic sampling, eurytopic haplotypes would likely appear even among Zaïrian populations, although all are separated by moderate to large rivers (not all of which are shown in Figure 3.6).

Five statistically-useful genetic measures of interpopulational distance were therefore calculated. The first is Nei, Tajima and Tateno's (1983) modified angular transformation genetic distance, which is given by:

$$D_A = \frac{1}{r} \sum_{j=1}^{r} \left(1 - \sum_{i=1}^{m_j} \sqrt{x_{ij} y_{ij}}\right)$$

where m_j is the number of alleles at the jth locus, r is the number of loci examined, and x_{ij} and y_{ij} are the frequencies of the ith allele in populations x and y, respectively. For the case of a single locus, D_A reduces to:

$$D_A = 1 - \sum_{i=1}^{m} \sqrt{x_i y_i}$$

D_A is useful because alleles with frequency zero in either population do not contribute to genetic distance. D_A thus quantifies genetic distance based on eurytopic haplotypes alone.

This measure's insensitivity to low-frequency alleles also minimizes bias introduced by different sample sizes among populations (Nei, Tajima and Tateno, 1983). Other similar angular transformation distance measures were tested (Cavalli-Sforza and Edwards, 1967; Nei, 1978; Reynolds, Weir and Cockerham, 1983). Results from these other distance measures were identical to those described for D_A, and therefore are not presented.

D_A quantifies genetic distance at an allelic level, which assumes that all haplotypes are equidistant. Genetic distance at the DNA sequence level is also informative, and can discern evolutionary events on a more distant time scale. Four measures of sequence-level genetic distance were therefore calculated. The first, mean pairwise sequence difference between populations, is given in Nei (1987) as:

$$\hat{d}_A = \hat{d}_{XY} - (\hat{d}_X + \hat{d}_Y)/2$$

where the average number of nucleotide substitutions between DNA haplotypes from populations x and y, and is given by:

$$\hat{d}_{XY} = \Sigma_{ij} \hat{x}_i \hat{y}_j d_{ij}$$

and the average number of nucleotide substitutions for a randomly chosen pair of haplotypes *within* a population (where n_x sequences have been sampled) is:

$$\hat{d}_X = \frac{n_X}{n_X - 1} \Sigma_{ij} \hat{x}_i \hat{x}_j d_{ij}$$

where x_i-hat and x_j-hat are the frequencies of the *i*th and *j*th haplotypes in population x, and d_{ij} is the number of nucleotide differences per site between the *i*th and *j*th haplotypes. Because the calculation of mean nucleotide difference corrects for intraspecific polymorphism, it is particularly useful for interpopulational comparisons within species (Nei, 1987).

Analogous measures of modal, minimum and maximum pairwise sequence difference were also calculated among populations. While necessarily correlated with mean

sequence difference, these measures have useful properties in their own right. Like D_A, minimum pairwise difference is sensitive to the effects of eurytopic haplotypes. Populations sharing a haplotype will be separated by a minimum pairwise distance of zero. Populations not linked by eurytopic haplotypes will be distant in proportion to the number of nucleotide differences between their most similar alleles. Maximum pairwise sequence difference has the opposite advantage: it ignores eurytopic haplotypes entirely. It is equivalent to measures of maximum pairwise sequence differences often used in studies of interspecific relationships (e.g. Ruvolo *et. al.*, 1994) because of their relevance to Tajima's neutral coalescent model (Tajima, 1983). Finally, modal pairwise sequence difference bears a close relationship to the critical parameters of "mismatch" and "intermatch" distribution analysis, described in Chapter 4 (Rogers and Harpending, 1992; Harpending *et. al.*, 1993), and is calculated for this reason.

Matrices of genetic distance among the 19 geographically-localized sampling locations were calculated for each distance measure. Each matrix was compared to a matrix of great-circle geographic distances among populations. These comparisons are depicted graphically in Figures 3.7 and 3.8. Although relationships are approximated by least-squares lines, individual points correspond to *pairwise* distances (geographic and genetic), and are thus statistically interdependent. Standard parametric analyses such as least-squares regression are therefore inappropriate. Instead, Mantel tests of matrix correlation were used to test the significance of each relationship (Mantel, 1967). A standardized form of the Mantel statistic, r, was calculated to facilitate comparison among analyses (Smouse, Long and Sokal, 1986). This statistic varies between -1 (perfect negative correlation) and +1 (perfect positive correlation) and is equivalent to the Pearson's correlation coefficient between the two matrices. Probabilities were calculated by a Monte-Carlo matrix permutation procedure (Hope, 1968), and were computed using 2000 permutations of the data matrices with the program "The R Package" (Legendre and Vaudour, 1991).

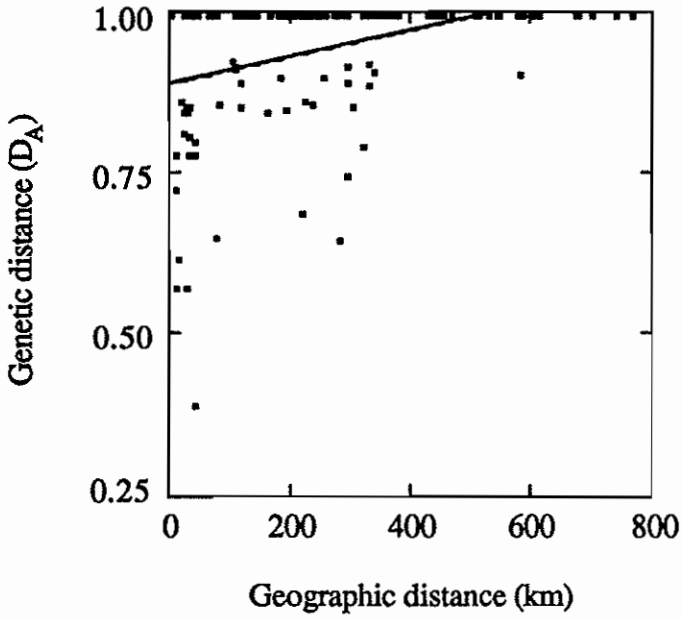


Figure 3.7: Relationship between Nei, Tajima and Tateno's (1983) modified angular transformation genetic distance (D_A) and pairwise geographic distance for 19 chimpanzee populations. D_A measures genetic distance between populations using allele frequencies.

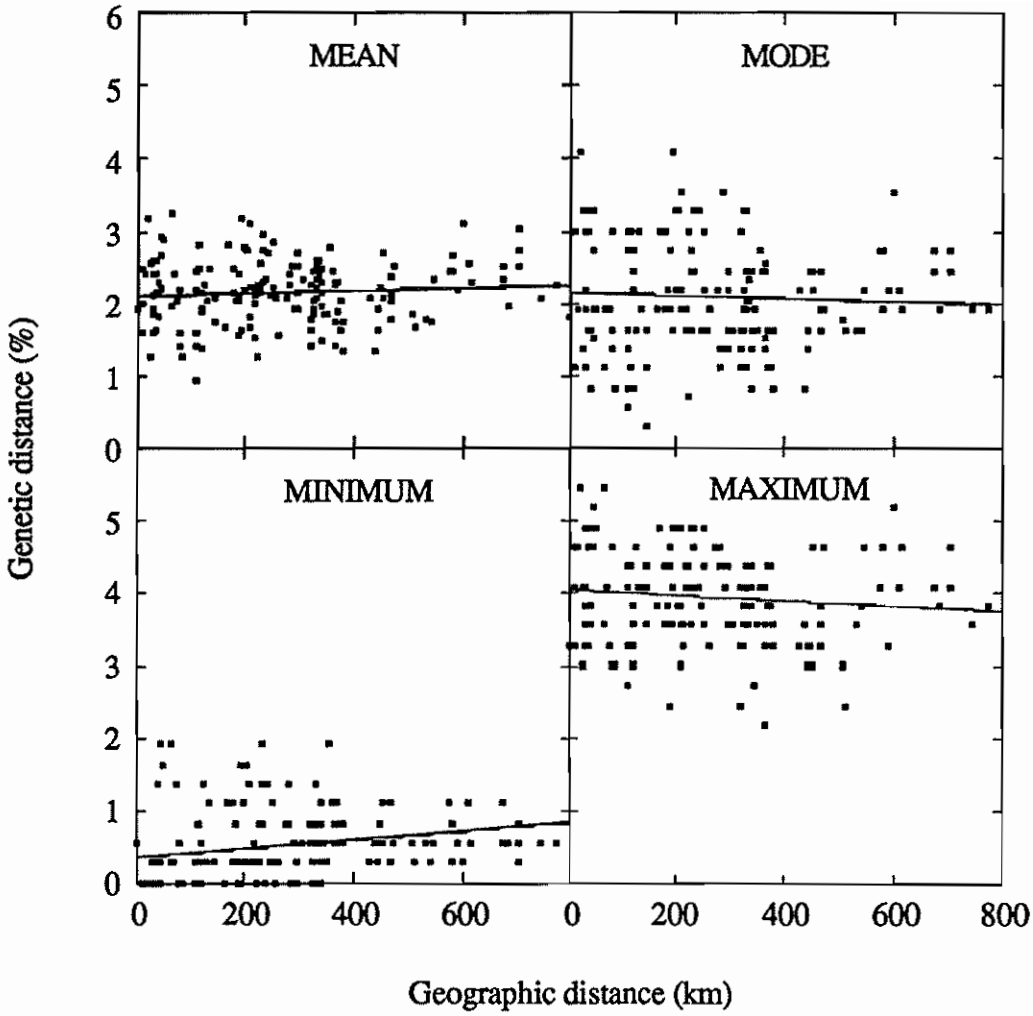


Figure 3.8: Relationships between pairwise genetic distance (at the DNA sequence level) and pairwise geographic distance for 19 chimpanzee populations. Genetic distance measures are mean, modal, minimum and maximum sequence differences between pairs of populations, expressed as percent nucleotide dissimilarity. Points are therefore not independent. Geographic distances are great-circle distances (km). Lines are fit using the least-squares method.

The relationship between D_A and geographic distance was positive and highly significant ($r = 0.37$; $p = 0.0005$). Maximum possible D_A values of 1 were, however, observed over the full range of geographic distances tested. The relationship is therefore both non-linear and heteroscedastic. When Eastern forests are considered separately, the relationship is still highly significant ($r = 0.40$; $p = 0.0005$). The analogous correlation for Zaïrian populations could not be calculated, since eurytopic haplotypes were absent among populations within Zaïre, making D_A among all Zaïrian populations 1. No significant correlation was observed for the relationship between mean, modal, maximum or minimum pairwise sequence difference among populations and geographic distance (mean: $r = 0.08$, $p = 0.33$; mode: $r = -0.05$, $p = 0.41$; minimum: $r = 0.24$, $p = 0.09$; maximum: $r = -0.02$, $p = 0.28$; see Figure 3.8). This relationship remained statistically insignificant when Eastern populations were considered separately (mean: $r = 0.11$, $p = 0.30$; mode: $r = -0.03$, $p = 0.49$; minimum: $r = 0.25$, $p = 0.09$; maximum: $r = -0.07$, $p = 0.37$), and when Zaïrian populations were considered separately (mean: $r = 0.13$, $p = 0.41$; mode: $r = 0.54$, $p = 0.08$; minimum: $r = 0.31$, $p = 0.22$; maximum: $r = 0.18$, $p = 0.39$).

The significant correlation between D_A and distance, and the lack of a significant correlation for any sequence-based measure, probably indicates a time-scale effect. Allelic measures such as D_A provide a window into evolutionary events which is more recent than that provided by DNA sequence-based measures. Drift and fixation of entire alleles is a faster evolutionary process than drift and fixation of point mutations (Nei, 1987; Li and Graur, 1991). If distance does influence the genetic structuring of eastern chimpanzee populations on the DNA sequence level, the trend is invisible to the analyses described above. This time-scale effect is consistent with the notion that forest expansion events of principal importance to the current distribution of eastern chimpanzees are late Pleistocene in origin.

Spatial autocorrelational analysis

Spatial autocorrelational analysis (Sokal and Oden, 1978a; Sokal and Oden, 1978b; Legendre, 1993) examines the strength of association between a variable and itself as a function of spatial distance. Since spatial distance can be subdivided into classes, this association can be examined across a range of distance classes, allowing the emergence of complex, non-linear patterns. Spatial autocorrelation has contributed significantly to the study of ecology, although the scope of its application is potentially much greater (Sokal and Oden, 1978a; Sokal and Oden, 1978b; Oden, 1984; Legendre, 1993). In particular, spatial autocorrelational techniques are useful for genetic analysis (Sokal and Oden, 1978a; Sokal and Oden, 1978b; Barbujani, 1987; Bertorelle and Barbujani, 1995), and have been applied extensively to the study of geography and human genetics (Sokal and Menozzi, 1982; Sokal, Harding and Oden, 1989; Barbujani *et. al.*, 1995).

Two commonly-used statistics to measure spatial autocorrelation are Moran's *I* (Moran, 1950) and Geary's *c* (Geary, 1954). Because of its convenient statistical properties (Sokal and Oden, 1978a; Cliff and Ord, 1981) and because of its history of application to human genetic data (Sokal and Menozzi, 1982; Sokal, Harding and Oden, 1989; Bertorelle and Barbujani, 1995), only the former measure will be used in the present study. For allele frequency data, Moran's *I* is given by:

$$I = \frac{n \sum_{ij} w_{ij} (p_i - \bar{p})(p_j - \bar{p})}{W \sum_{i=1}^{n} (p_i - \bar{p})^2}$$

where p_i and p_j are the frequencies of allele p in populations i and j, respectively, and p-bar is the mean over all n populations (Barbujani, 1987). The variable w_{ij} represents an element in a matrix of geographic connectedness for a specific distance class. It is binary, and takes a value of one if the *i*th and *j*th localities are separated by a geographic distance which falls into the distance class being considered, or zero if they are not. *W* is the sum

of all w_{ij} within the distance class. Moran's *I* varies between -1 (perfect negative autocorrelation) and +1 (perfect positive autocorrelation) for very large sample sizes.

Bertorelle and Barbujani (1995) have proposed an analogous statistic, *II*, for the analysis of genetic data at the DNA sequence level. This "AIDA" (autocorrelation index for DNA analysis) is given by:

$$II = \frac{n \sum_{i=1}^{n-1} \sum_{j>i}^{n} w_{ij} \sum_{k=1}^{S} (p_{ik} - \bar{p}_k)(p_{jk} - \bar{p}_k)}{W \sum_{i=1}^{n} \sum_{k=1}^{S} (p_{ik} - \bar{p}_k)^2}$$

where *p* now becomes a binary variable representing the identity (1) or nonidentity (0) of a nucleotide (or restriction site) along a vector of *S* sites representing one of n individual sequences in a population sample (Bertorelle and Barbujani, 1995). *II* has the advantage that, because it treats each DNA sequence as an independent observation, it gains considerable statistical power over Moran's *I*, which treats each population as an independent observation. Also, *II* can be calculated among individuals within a single population, providing a value at the zero distance class.

A plot of autocorrelational strength as a function of geographic distance is called a spatial autocorrelogram (Sokal and Oden, 1978a). Autocorrelograms which decrease monotonically from positive to negative with distance indicate a clinal trend; those showing other, more complex patterns (e.g. localized minima and maxima), may indicate varying kinds and degrees of geographic patchiness (Sokal, Harding and Oden, 1989; Bertorelle and Barbujani, 1995). Since autocorrelational analysis by nature assumes that data are interdependent, parametric techniques are inadequate for assessing the statistical significance of a spatial autocorrelogram (Oden, 1984; Legendre and Fortin, 1989). Rather, statistical significance should be inferred by comparison of the observed autocorrelation index to a null distribution generated by a large number of random

permutations of the data. In this case, a Bonferroni correction should be used in which the significance level α at which the null hypothesis of no spatial autocorrelation is rejected is adjusted to $\alpha'=\alpha/n$, where n is the number of simultaneous statistical tests (distance classes; Oden, 1984).

With the allele frequency data given in Table 3.5, Moran's *I* was calculated for each of the 24 eurytopic alleles identified in the study using the computer program "The R Package" (Legendre and Vaudour, 1991). Nine distance classes of equal width (86 km) were defined. To assess the statistical significance of each of the resulting 24 autocorrelograms, a null distribution of *I* was generated for each autocorrelogram at each distance class, using 1000 random permutations of the data. Four of the resulting spatial autocorrelograms were significant by the Bonferroni criterion. The autocorrelogram in Figure 3.9 shows the average value of *I* at each distance class for the four Bonferroni-significant autocorrelograms, following the methodology of Sokal, Harding and Oden (1989). The level of statistical significance is indicated by asterisks, and is calculated as the mean statistical probability associated with each distance class.

The jaggedness of the resulting autocorrelogram probably results from the small number of alleles (four) and the small number of genetic systems (one) on which the analysis is based. The associated sampling error makes it impossible to classify the autocorrelogram unambiguously as "clinal" or "patchy" (Sokal, Harding and Oden, 1989). Nevertheless, positive autocorrelation is observed only for the first distance class. The fact that all other distance classes show negative autocorrelation suggests that substantial gene flow in eastern chimpanzees is limited to distances less than approximately 100 km. The observation that Moran's *I* does not decrease monotonically after the third distance class suggests a pattern either of "local patchiness" or "long-distance differentiation," rather than a strict cline (Sokal, Harding and Oden, 1989; Barbujani *et. al.*, 1994).

To investigate this relationship further, spatial autocorrelational analyses were performed at the DNA sequence level using the computer program AIDA (Bertorelle and

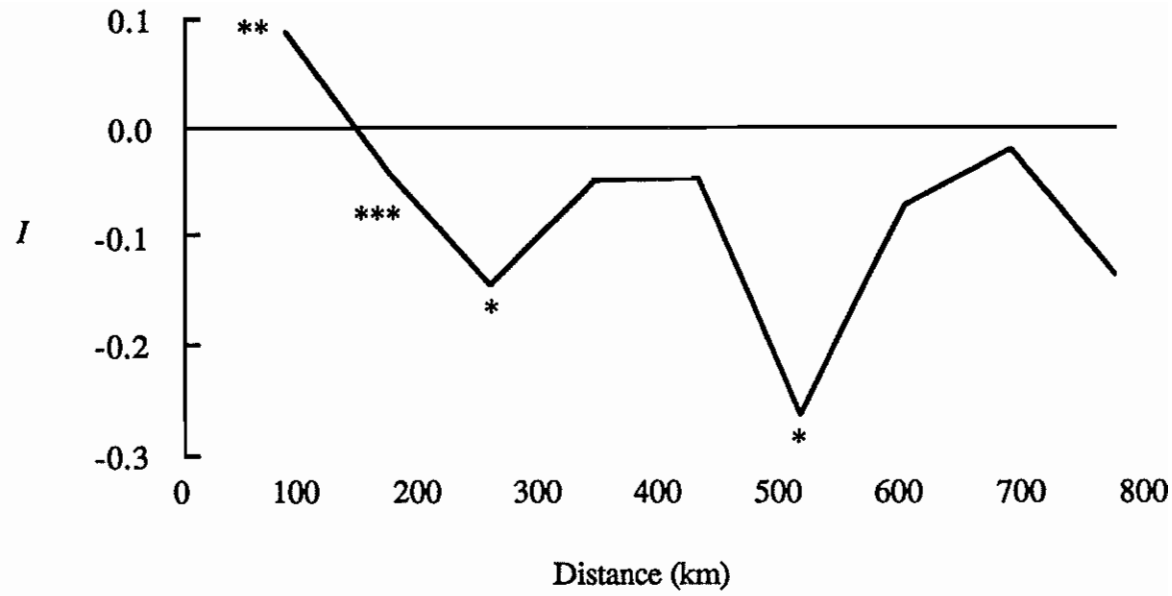


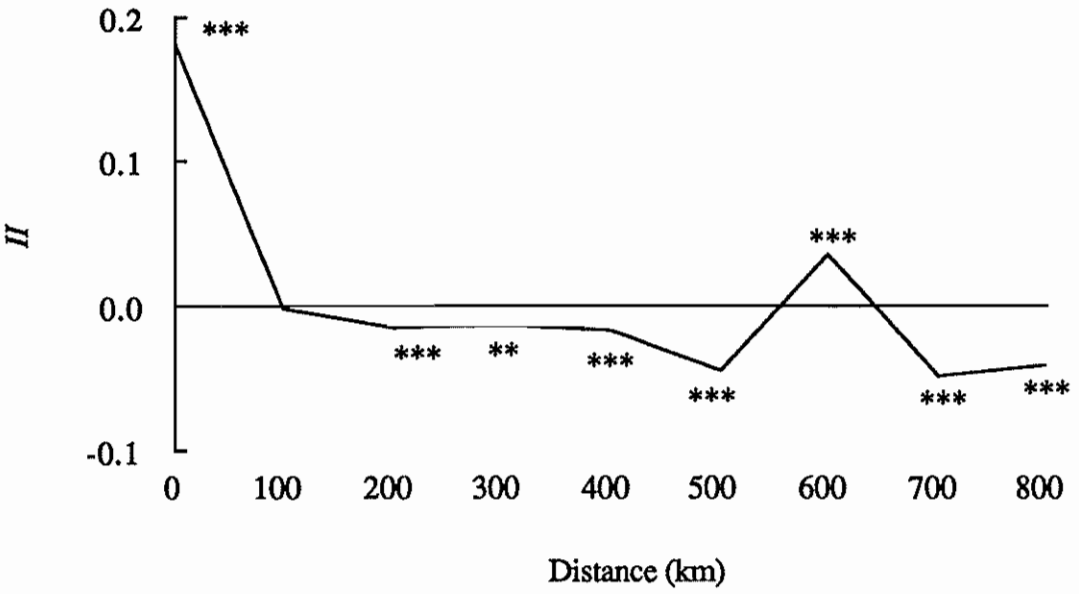
Figure 3.9: Spatial autocorrelogram based on allele frequencies. Values of *I* are averages of the four Bonferroni-significant autocorrelograms which resulted from autocorrelational analysis of 24 eurytopic alleles. Asterisks indicate the level of statistical significance, based on 1000 random permutations of the data (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$).

Barbujani, 1995). DNA sequences were coded as binary vectors according to the method of Bertorelle and Barbujani (1995). Analyses were performed on nine distance classes (including zero) of even width (100 km) using great-circle geographic distances. Statistical significance was assessed for each distance class using 1000 random permutations of the data. Figure 3.10 (top panel) shows the results of this analysis for the full data set of 255 sequences. The autocorrelogram is significant by the Bonferroni criterion, and is markedly positive at distance class zero.

Autocorrelation at the zero distance class may be artifactually high due to multiple representations of haplotypes within sampling locations. Such bias would be introduced if individual chimpanzees were accidentally sampled more than once (see Chapter 2). To account for the influence of double-sampling, a reduced data set was created in which identical haplotypes within sampling locations were assumed to represent double-sampled individuals. All identical haplotypes within sampling locations were collapsed, and the AIDA analysis was repeated on the resulting data set of 164 sequences. Results are presented in Figure 3.10 (bottom panel).

The general pattern observed for the reduced data set is fundamentally the same as that for the full data set. Both analyses show significant positive spatial autocorrelation at the zero distance class, and generally negative and decreasing autocorrelation at larger distance classes. The local increase at the 600 km distance class observed in the autocorrelogram for the full data set clearly results from the presence of identical haplotypes within sampling locations. Both AIDA analyses conform best to a pattern of long-distance differentiation (Sokal, Harding and Oden, 1989; Barbujani *et. al.*, 1994; Bertorelle and Barbujani, 1995) and are remarkably similar to the pattern observed for Italian humans over roughly equivalent geographic distances (Barbujani *et. al.*, 1995). This basic pattern remains consistent for AIDA analyses performed on Eastern and Zaïrian forests separately (Figure 3.11).

Full data set



Reduced data set

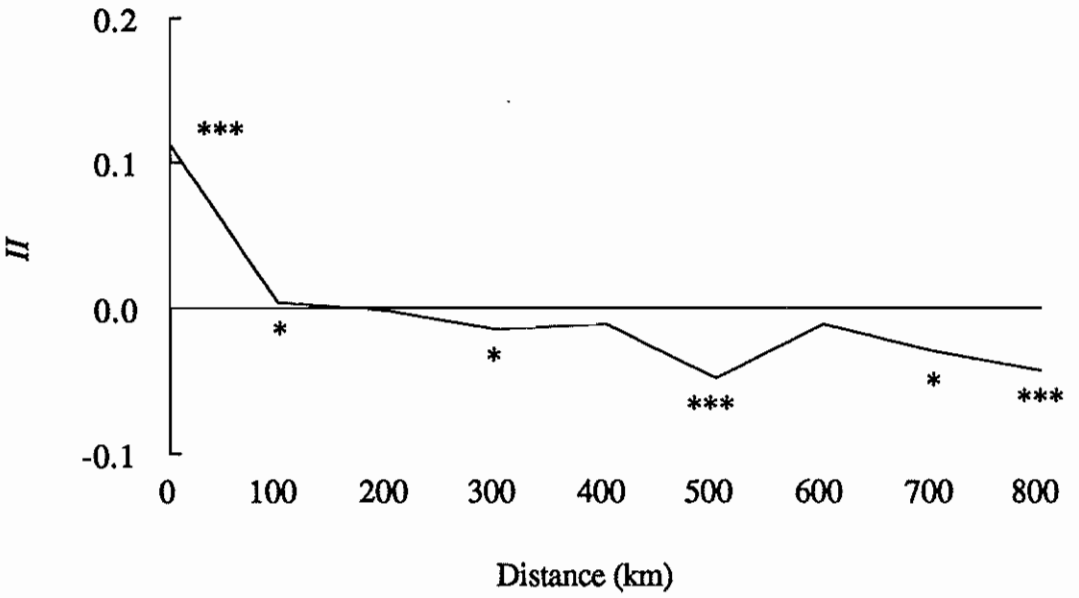


Figure 3.10: Spatial autocorrelograms based on DNA sequence data (AIDAs). The top panel shows the autocorrelogram for the full data set (255 sequences) and the bottom panel shows the autocorrelogram for a reduced data set (164 sequences) created by collapsing identical haplotypes within sampling locations. Asterisks indicate the level of statistical significance for each distance class, based on 1000 random permutations of the data. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$. Data were analyzed using the computer program AIDA (Bertorelle and Barbujani, 1994).

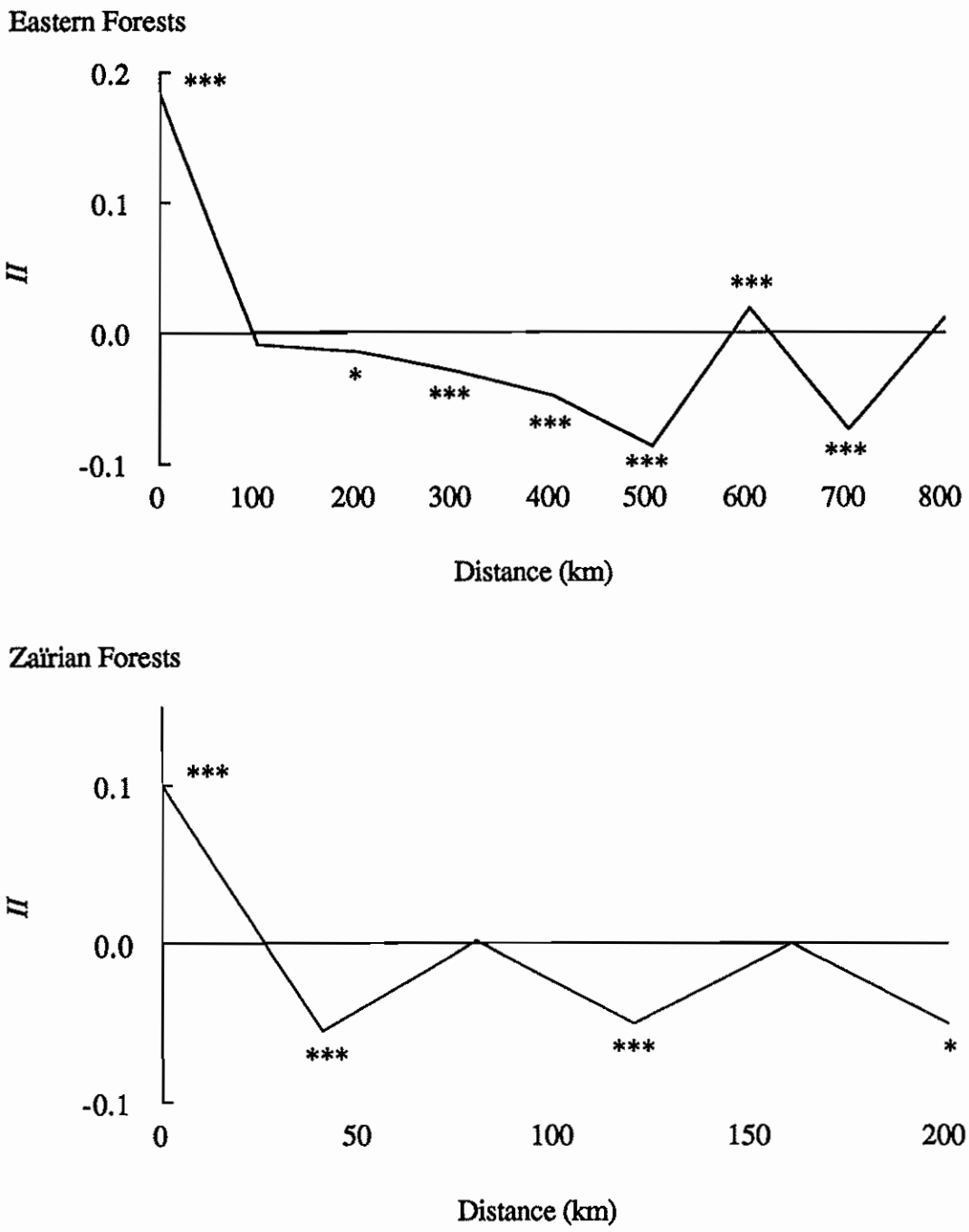


Figure 3.11: AIDAs for Eastern forests (top panel) and Zaïrian forests (bottom panel) separately. The Eastern forest data set consisted of 190 sequences and the Zaïrian forest data set consisted of 65 sequences. Asterisks indicate the level of statistical significance for each distance class, based on 1000 random permutations of the data (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$). Data were analyzed using the computer program AIDA (Bertorelle and Barbujani, 1994).

AIDA analysis was performed on a reduced data set in which eurytopic haplotypes were eliminated entirely. Within sampling locations, redundant endemic haplotypes were also collapsed to eliminate bias caused by high-frequency alleles. The resulting autocorrelogram (based on 88 sequences; Figure 3.12) retains the general properties of the autocorrelograms previously described. Even with the elimination of all identical and shared haplotypes in the sample, nucleotide sequences tended to be similar over short distances and different over long distances. This result demonstrates that sequence-level molecular evolutionary processes are generating unique, spatially-localized alleles in eastern chimpanzees. The general negativity of all AIDAs at distance classes greater than zero confirms the result based on allele frequency data that dispersion of alleles is restricted after distances of approximately 100 km. In reality, 100 km is a liberal estimate of the maximum distance over which DNA sequences are spatially autocorrelated. Analyses using greater numbers of smaller distance classes suggest that AIDAs become negative after distances as small as 60 km.

Conclusions

In general, the insular biogeography model of post-glacial dispersal from forest refugia does not have much explanatory power for east African chimpanzee genetic diversity. Forest area was a poor predictor of genetic diversity, accounting for a maximum of 30% of the variation in genetic diversity among populations by any measure. Distance from the "source" was an equally poor predictor. Within the study area, distance from any geographic point explained a maximum of only 37% of the variation in genetic diversity among populations in the expected direction. This is not a significantly higher proportion of the variation than is explained by present degrees of forest isolation, which explains a maximum of 34% of the variation in genetic diversity among populations. If eastern chimpanzee populations do retain a genetic signature of dispersal from any single refugium

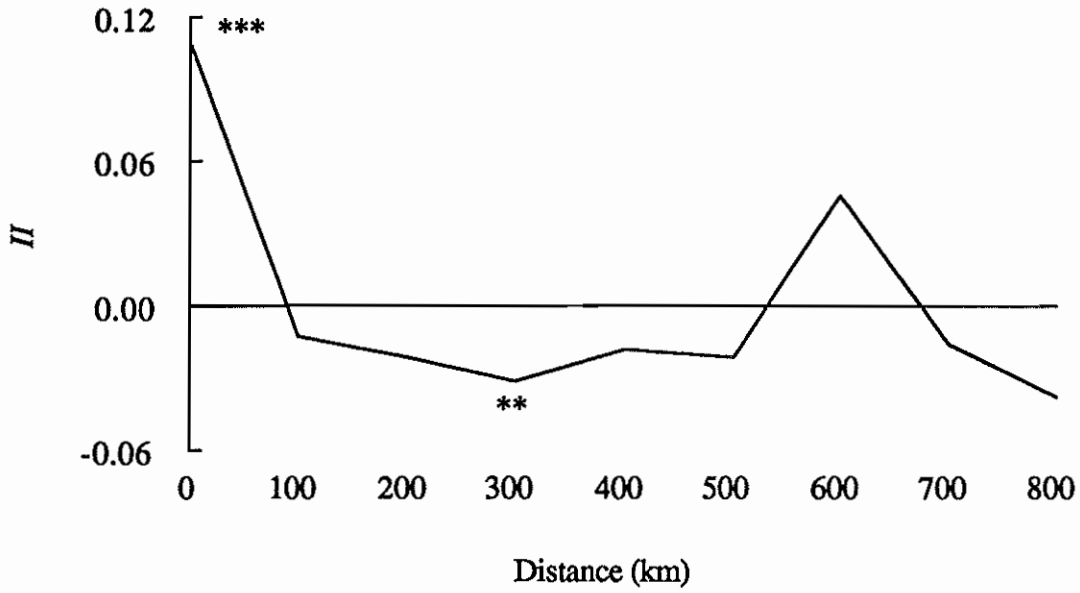


Figure 3.12: AIDA for endemic haplotypes. Endemic haplotypes were defined as those haplotypes found exclusively within a single sampling location. Redundant haplotypes within each sampling location were also collapsed. The autocorellogram is based on 88 sequences. Asterisks indicate the level of statistical significance for each distance class, based on 1000 random permutations of the data. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$. Data were analyzed using the computer program AIDA (Bertorelle and Barbujani, 1994).

within the study area, the evidence has been largely obscured.

Three hypotheses might account for the lack of the biogeographic model's explanatory power. The first is that eastern chimpanzee population history has been characterized by patterns of geographic expansion more complex than the one tested here. Multiple refugia and complex dispersal routes (Kingdon, 1981; Colyn, 1991) would make the prediction of population genetic diversity difficult because of the complexity of the biogeographic model involved. A hypothesis of simple expansion may also be an oversimplification. Given the temporal instability of the East African Pleistocene climate (Moreau, 1963; Cerling, Hay and O'Neil, 1977), secondary, tertiary and higher-order expansion and contraction events are possible for the forests in which eastern chimpanzees have lived.

The second (and not mutually-exclusive) hypothesis is that subsequent biological events have obscured any genetic record of dispersal from refugia. Even if the initial diversity of chimpanzee populations were perfectly predicted by area and distance parameters, subsequent gene flow could have overwritten this pattern. However, gene flow has clearly been limited by distance, as indicated by the positive correlation between geographic and genetic distance at an allelic level, and by spatial autocorrelational analysis. Dispersal barriers other than geographic distance also probably contribute to genetic differentiation among populations. Large rivers in Zaïre, for example, further limit gene flow, as evident in the higher haplotype endemicity within Zaïrian populations than within eastern populations. Autocorrelational analyses at the allelic and nucleotide level suggests a maximum distance of approximately 100 km over which gene flow has significantly impacted the genetic structure of the subspecies.

The third hypothesis to explain the insular biogeography model's lack of predictive power is that chimpanzees were never restricted to Pleistocene forest refugia. Chimpanzees are an extraordinarily vagile species, showing impressive ranging abilities and ecological flexibility (Kortlandt, 1983; Goodall, 1986). It is possible that, given these abilities,

chimpanzees were able to live outside of refugia, as well as in them, during periods of minimal global temperature (see Chapter 1). Chimpanzees may, for example, have ranged throughout an ecological mosaic of savannah, woodland, and gallery forest during times when other, more "obligate" forest taxa, were restricted to refuge forests. If so, then extensive migration would have occurred throughout the history of the subspecies. Indeed, AMOVA analysis supports this contention by demonstrating only a small degree of population subdivisioning within the subspecies, regardless of how populations are defined, and high rates of interpopulation migration.

Comparisons with humans

One striking feature of the analyses described above is the general similarity of eastern chimpanzees to humans. Overall levels of genetic variability in eastern chimpanzees are comparable to, and even slightly lower than, those in humans. Variability estimates for eastern chimpanzees will probably increase with additional sampling, but analogous estimates in humans are not likely to do so (Ruvolo *et. al.*, 1993; Ruvolo *et. al.*, 1994). Therefore, differences in genetic variability between the two taxa are likely to change as research continues. Eastern chimpanzees differ from humans principally in their relative lack of inter-regional mitochondrial variation (even using the smallest human estimate from the literature), and in their correspondingly greater proportion of among-population variation. This difference probably results from order-of-magnitude differences in the geographic scale over which the two taxa range. Genetic differentiation at the nucleotide level in eastern chimpanzees is comparable to that in humans when viewed on a similar geographic scale; AIDA analyses of Italian humans over a scale of 800 km (equivalent to the scale over which chimpanzees were examined in the present study) showed a clinal pattern indistinguishable from that shown for eastern chimpanzees (Barbujani *et. al.*, 1995).

The similarities between humans and eastern chimpanzees are particularly intriguing considering that humans are generally considered unusual in their genetic homogeneity. Early analyses of blood group and protein loci (Lewontin, 1972; Nei and Roychoudhury, 1974; Latter, 1980) documented levels of inter-racial genetic variation in humans which accounted for only a small percentage of the total genetic variation within the species. Lewontin (1972) interpreted low genetic variation among human races as "surprising" principally because of *a priori* expectations based on morphological differences between races, the magnitude of which were often exaggerated due to biases in human perception. Later analyses of human mitochondrial DNA reconfirmed the observation that inter-regional variation in humans was low (Cann, Stoneking and Wilson, 1987; Stoneking *et. al.*, 1990; Merriwether *et. al.*, 1991; Vigilant *et. al.*, 1991; Ward *et. al.*, 1991; Ruvolo *et. al.*, 1993). The regional divisioning found in these studies was, however, slightly higher than previous estimates based on more slowly-evolving nuclear genetic systems, and considerably less than the mitochondrial estimates given by Excoffier, Smouse and Quattro (1992), as noted earlier.

Comparative studies of mitochondrial DNA between humans and apes similarly supported the notion that genetic diversity in humans was low (Brown, 1980; Ferris, Wilson and Brown, 1981; Ruvolo *et. al.*, 1994). This observation had significant implications for the study of anthropological genetics, forming an important basis for the "mitochondrial Eve" hypothesis of modern human origins (Cann, Stoneking and Wilson, 1987; Vigilant *et. al.*, 1991). In this context, low mitochondrial genetic diversity in humans was initially "surprising" because it implied a common human ancestry too recent to be compatible with archaeological expectations (Wolpoff, 1989; Ruvolo *et. al.*, 1993).

Human genetic diversity is undoubtedly low. However, to say that human genetic diversity is "different" in kind or degree requires direct comparative data from closely-related taxa of an equivalent nature. This study suggests that many "uniquely human"

genetic characteristics may be artifacts of comparisons made across taxonomic boundaries (Brown, 1980; Ferris, Wilson and Brown, 1981; Ruano *et. al.*, 1992; Ruvolo *et. al.*, 1994). Systematic geographic analyses of genetic variability within the remaining great ape subspecies may someday indicate that human genetic variability is more typical in its extent and apportionment than previously thought, except for the unusually large geographic range over which it is distributed.

Chapter 4: Mismatch distribution analysis

Mismatch distribution theory

Measures of population genetic diversity such as those discussed in the previous chapter are useful summary statistics for describing DNA sequence variation within a population. However, because they collapse all pairwise comparisons between DNA sequences into a single value, such point estimators ignore variation contained within the sample. Pairwise genetic distances among individuals are more accurately represented as a *distribution* of values. A distribution of pairwise genetic differences between individuals within a population is termed a "mismatch distribution" (Harpending *et. al.*, 1993), and is depicted graphically as a frequency distribution, often normalized to unit area.

Mismatch distribution analysis complements tree-based methods of mitochondrial DNA analysis in attempting to reconstruct population history. Although the "ultimate currency" of a historical molecular study is in many ways an accurate phylogenetic tree, such trees are notoriously difficult to reconstruct. Mismatch distribution analysis of DNA sequence data (or restriction site data) has the advantage that its inferences do not rely on accurate topological reconstructions of phylogeny. Figure 4.1, redrawn from Rogers and Jorde (1995) demonstrates why this property is particularly advantageous for populations which have undergone sudden expansion. In expanding populations, the time-window during which phylogenetically-informative mutations might arise is discouragingly narrow. Inferences from the topologies of reconstructed phylogenetic trees should not therefore be expected to yield accurate information about population history for populations which have undergone sudden expansion. Mismatch distribution analysis is specifically designed to estimate the times at which populations have undergone expansion after their origin, as described below.

As Figure 4.1 shows, a mismatch distribution may be viewed as "shorthand

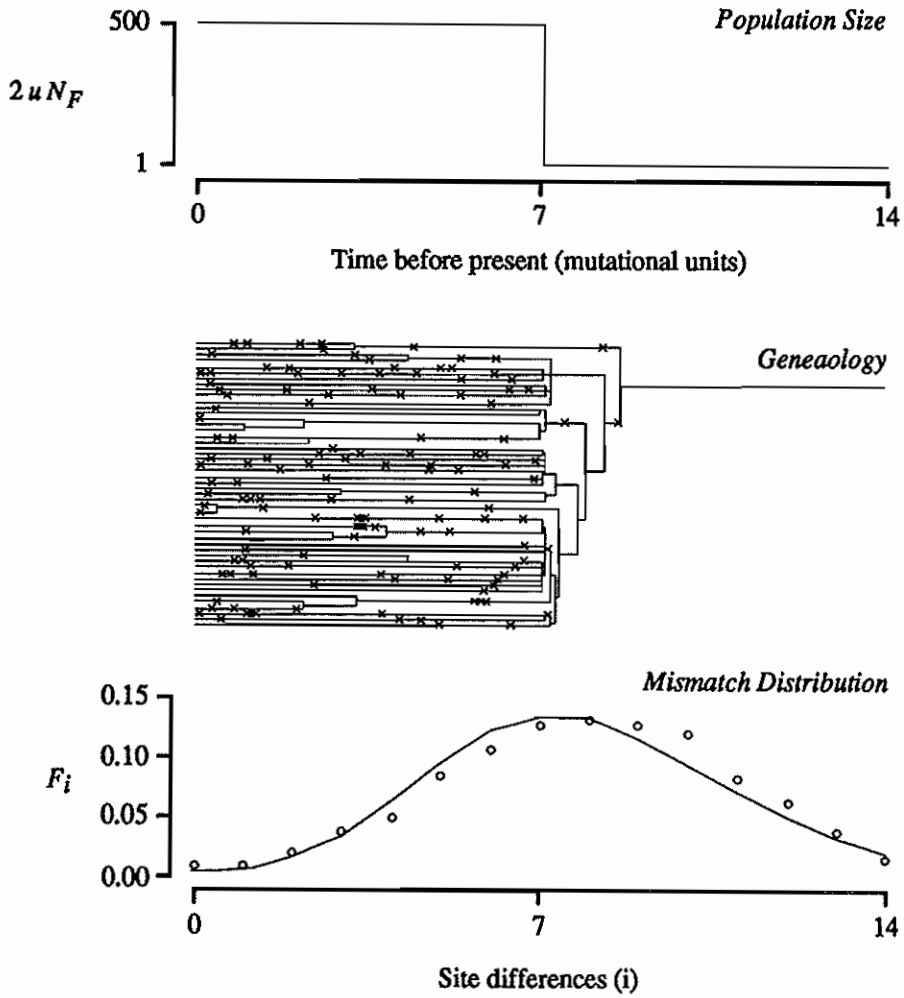


Figure 4.1: Relationships between population history, phylogeny and the mismatch distribution (redrawn from Rogers, 1995b). The top panel shows a population history of sudden expansion at 7 mutational time units before the present. The middle panel shows the corresponding phylogeny, with crosses representing mutations. Note that relatively few mutations occur during the time of population expansion. The lower panel shows the mismatch distribution for these data (open circles represent empirical values; the solid line represents the theoretical mismatch distribution calculated for this population).

notation" for a phylogeny (see below). Mismatch distribution analysis implicitly recognizes that information about a population's history is contained throughout the phylogeny, and not just in the relationships among the most divergent lineages or major clades (Rogers and Harpending, 1992). In focusing on modal pairwise genetic differences (rather than maximum differences), mismatch distribution analysis therefore provides a more recent time window into evolutionary events than do tree-based analyses. Unfortunately, the mathematical and statistical properties of mismatch distributions are complicated. Because each value in the distribution is calculated as the genetic distance between an individual and *all other* individuals within the population, the values of the distribution are not statistically independent. As a result, it is inappropriate to describe a mismatch distribution with reference to any standard parametric model (Di Rienzo and Wilson, 1991; Slatkin and Hudson, 1991). Furthermore, the DNA sequences themselves are non-independent in that they are related genealogically. A body of research is currently emerging which explores the statistical properties of mismatch distributions given these limitations (Rogers and Harpending, 1992; Rogers, 1995a; Rogers and Jorde, 1995).

Mathematical properties of mismatch distributions

The shape of a mismatch distribution reflects the recent demographic history of the population from which the genetic data were sampled. As described by Slatkin and Hudson (1991) and Rogers and Harpending (1992), mismatch distributions derived from populations which have undergone recent expansions show characteristic wave-like form. This contrasts with populations which have been demographically stable for long periods of time (equilibrium populations), which show erratic distributions, both under computer simulation and with real data (Slatkin and Hudson, 1991; Rogers and Harpending, 1992; Harpending *et. al.*, 1993). Rogers and Harpending (1992) define three parameters which delimit the shape of a wave-like mismatch distribution. They derive these parameters from

a continuous approximation of the theoretical mismatch distribution, itself derived from population genetic theory (Rogers and Harpending, 1992), under a simplified demographic model of "sudden expansion." This approach proves surprisingly robust to violations of its underlying assumptions (Rogers, 1992; Rogers and Harpending, 1992; Rogers and Jorde, 1995).

The three parameters defined by Rogers and Harpending (1992) are:

4.1) $\theta_0 = 2N_0u$

4.2) $\theta_1 = 2N_1u$

4.3) $\tau = 2ut$

where N_0 is the population size (of breeding females) before population expansion, N_1 is the population size after expansion, and t is the time since population expansion occurred. u is "the per-generation probability that a mutation strikes some nucleotide in the region under study" (Rogers and Harpending, 1992), and is therefore specific to the gene sequence used, and its length. Also:

4.4) $u = m_T\mu$

where m_T is the length (in nucleotides) of the gene region, and μ is the number of nucleotide changes per site per generation. The shape of a mismatch distribution is defined by the values of these parameters. τ defines the location of the crest of the wave, and is directly proportional to the time at which the population underwent expansion:

4.5) $\tau = 2m_T\mu t$, or $t = \tau/2m_T\mu$

where t is the time of expansion in generations. In absolute years, the time since expansion, T, is therefore:

4.6) $T_{(abs.\ yrs.)} = G(\tau/2m_T\mu)$, or $G(t)$

where G is the generation time (years per generation).

In a series of mismatch distributions taken from a population undergoing sudden expansion, the crest of the wave moves to the right (τ increases) as time since the population expansion event increases (t increases). Figure 4.2 (redrawn from Rogers and Harpending, 1992) demonstrates this relationship. The time since population expansion, t, can therefore be calculated as $\tau/2u$, if τ is experimentally determined (see below), and if u is assumed to be constant with time (i.e. under the "infinite sites" model of mutation on which the model is based). θ_0 defines the slope of the leading (right-hand) face of the wave, and, like τ, can be estimated empirically from genetic data. The size of the initial population can be calculated as $\theta_0/2u$. Finally, θ_1 defines the value of the wave's y-intercept. An empirical estimate of θ_1 allows N_1, the size of the population after expansion, to be calculated roughly.

The mathematical theory of mismatch distributions is derived from a population genetic model which "refers to an average over an infinite ensemble of realizations of the evolutionary process" (Rogers and Harpending, 1992). In other words, a mismatch distribution with accurate values of θ_0, θ_1 and τ would only be expected if a researcher could sample a population many times after many independent realizations of the evolutionary process. This is clearly impossible in the real world, since a population can experience only a single history. The relationship between the theoretical values of θ_0, θ_1 and τ and their empirically derived counterparts is therefore unexpected, since, for real data, the sample size is always one. Empirically-derived parameter values should thus be viewed with appropriate caution. Nevertheless, simulation studies suggest that empirical

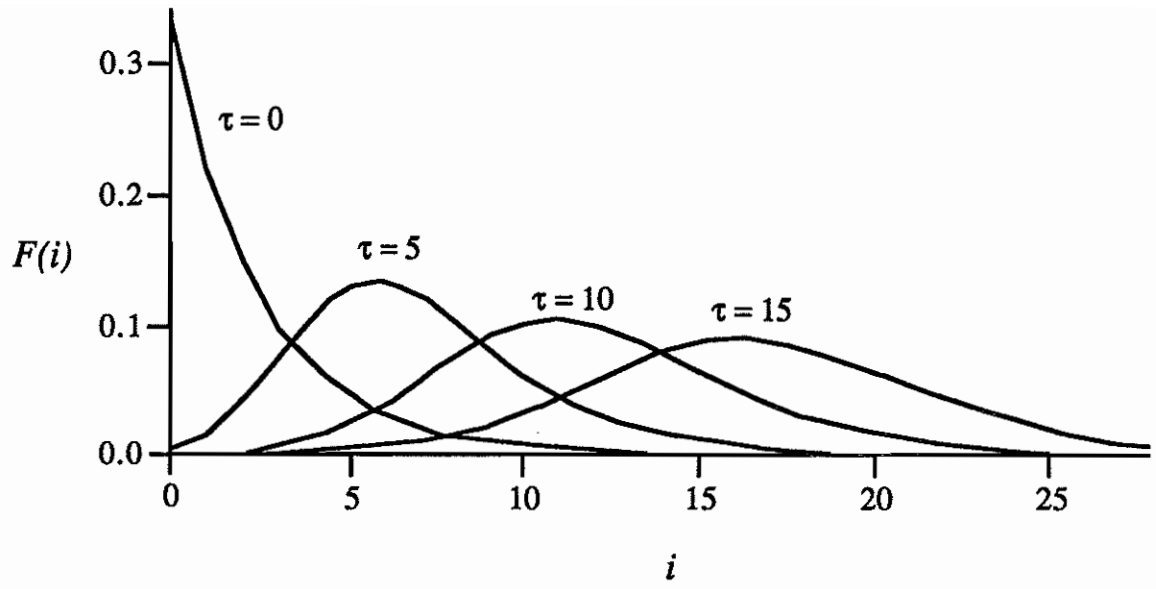


Figure 4.2: Relationship between τ and the shape of the mismatch distribution (redrawn from Rogers and Harpending, 1992). τ defines the location of the crest of the wave. As the time since population expansion increases, the crest of the wave moves to the right. The line for $\tau = 0$ represents the theoretical mismatch distribution for a population which has remained constant in size for a long time (an equilibrium population). i is the absolute number of pairwise nucleotide differences.

values of θ_0, θ_1 and τ are, in practice, legitimate estimators of N_0, N_1 and t respectively for populations which have undergone expansions (Rogers and Harpending, 1992; Harpending *et. al.*, 1993; Rogers, 1995a), and perhaps even for populations which show non-wave-like mismatch distributions (Rogers, 1995c).

Statistical properties of mismatch distributions

Given a wave-like mismatch distribution, what form of population growth is consistent with the data? Two techniques have emerged to answer this question. The first, described by Harpending *et. al.* (1993) is based on the observation that mismatch distributions from expanding populations tend to be smoother than distributions generated from equilibrium populations. Harpending *et. al.* (1993) propose an *ad hoc* statistic, r, which is calculated as the sum of squared differences between sequential ordinal values in a mismatch distribution. Low values of r indicate a smooth distribution and a population history characterized by recent expansion; high values indicate a rough distribution and a population history characterized by demographic stability.

Computer simulation of mismatch distributions demonstrates that r does discriminate between populations which have undergone recent expansion and those which have been demographically stable (Harpending *et. al.*, 1993). In one simulation study, mismatch distributions from populations which have experienced 10^3-fold population growth show r values of 0.012 (standard deviation of 0.006), while simulated distributions from equilibrium populations show r values of 0.26 (standard deviation of 0.20; see Harpending *et. al.*, 1993, Figure 2). These values tend to be representative of expanded and equilibrium populations in general, both in magnitude and in the fact that values for each type of population tend not to overlap significantly (see also Sherry *et. al.*, 1994, Figure 5). The precise values which r takes will vary depending on the size of the sample, the length of the DNA sequence under study, and the type of data (e.g. DNA

sequence data versus restriction site data). The *r* statistic is also resilient to the effects of population subdivisioning, differing levels of migration between divided subpopulations, low-resolution data, and small sample size (Harpending *et. al.*, 1993; Sherry *et. al.*, 1994).

The second approach to estimating population growth on the basis of a mismatch distribution has been pioneered by Rogers (Rogers, 1995a; Rogers and Jorde, 1995), and involves the construction of confidence regions around θ_0 and τ. In this approach, a collection of simulated data sets (of the same size as the data set under investigation) is generated using a range of values of θ_0 and τ. Data sets of this type are simulated across a range of hypothetical population histories involving differing degrees of population growth. The number of simulated data sets for which the simulated parameter values θ_0 and τ are at least as large as their empirically-derived counterparts is used to define a range of acceptance for these variables. This range of acceptance represents a 95% confidence region, depicted visually as a series of graphs, one for each level of population expansion, in which closed circles represent values of θ_0 and τ which fall within the 95% limits (see Rogers, 1995a/b).

This method has been used successfully to falsify hypotheses about modern human origins (Rogers, 1995a; Rogers and Jorde, 1995). Rogers has generated confidence regions for human mitochondrial data which exclude scenarios of less than 10^3 fold population expansion. In other words, values of θ_0 and τ which are compatible with the empirical data are only observed under population histories in which the ratio of θ_1 to θ_0 is at least 1000. θ_0 may be used retrospectively to estimate θ_1, the empirical value of which is itself a poor estimator of the true parameter value (Rogers, 1995b). 1000-fold growth would imply an initial size for the human founding population too small to have been spread across three continents, as the multiregional model of modern human origins would predict (Rogers, 1995a; Rogers and Jorde, 1995).

The advantage of this technique is that it makes no distributional assumptions that are violated by the inherent non-independence of the data. Confidence intervals constructed in this way tend also to be resilient to roughness in the empirical mismatch distribution (Rogers, 1995c). The disadvantage of this technique (aside from computational intensity) is that, because of the broad range of potential population histories that can be evaluated, one can never thoroughly explore the full landscape of possible scenarios of population growth. The parameters used to define the population history under which confidence regions are generated are entirely at the discretion of the researcher. The hypothetical population history can thus contain any number of population expansions, contractions, and bottlenecks, as well as any conceivable level of subdivisioning and migration. This technique should not therefore be used for exploratory data analysis. It should, rather, be used in its narrowest possible application, to falsify hypotheses about population history that are derived independently.

Eastern chimpanzee mismatch distributions

Mismatch distributions for chimpanzee data were calculated using the 262 DNA sequences generated during this study (see Chapter 2 and Appendix 1). Since DNA sequences from Gombe (Morin *et. al.*, 1994a) contained large gaps of missing data, GEKA samples were not included in the analyses, except where specifically noted. Data from the Entebbe Zoo population (EEZO) were included in the analysis under the assumption that all Entebbe Zoo chimps were, in fact, *P. t. schweinfurthii* (this assumption is supported by phylogenetic analysis; see Chapter 5 for details). Entebbe Zoo chimps were, however, excluded from all geographically-based analyses since the geographic origins of these chimpanzees were undetermined. Distributions were generated and analyzed by the computer package "Mismatch: Computer Programs for Analysis of Mismatch Distributions," version 2b, available via ftp at anthro.utah.edu (Rogers, 1995b).

Figure 4.3 shows the mismatch distribution for eastern chimpanzees calculated

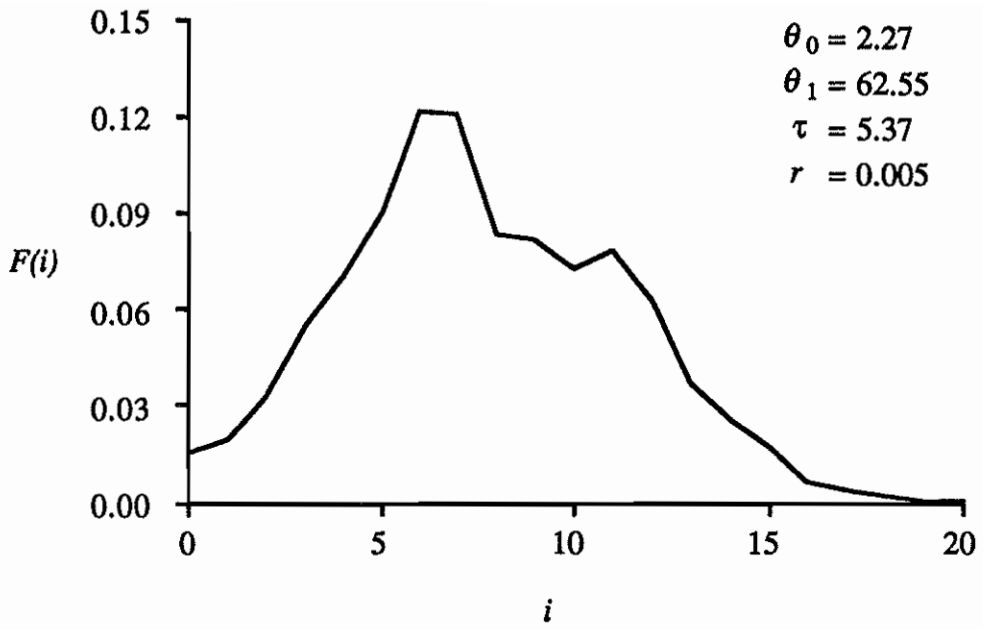


Figure 4.3: Mismatch distribution for 262 eastern chimpanzee d-loop subregion sequences generated in this study.

from 262 complete sequences. The distribution is wave-like in form, with a roughness value of $r = 0.005$. This small r value, the large sample size of sequences, and the large number of segregating sites (90) support the conclusion that the chimpanzee population in eastern Africa has undergone considerable recent demographic expansion. The empirical values of θ_0 and r are comparable to those described for human populations by Sherry *et. al.* (1994) using data from the human control region (hypervariable region 1, homologous to the sequence analyzed for chimpanzees in the present study). However, the empirical value of τ for the chimpanzee distribution (5.37) is slightly higher than the value reported in Sherry *et. al.* (1994) for the world human sample (4.26), and is outside the range of values given for individual human populations (confidence intervals around these estimates are considered below). The chimpanzee mismatch distribution shown in Figure 4.3 also differs from typical human mismatch distributions in that it shows some evidence of bimodality (Marjoram and Donnelly, 1994). Specifically, the chimpanzee distribution does not decrease monotonically from $i = 5$ to $i = 20$, but shows a slight local increase at $i = 11$. This pattern is explored more fully below.

As described in Chapter 2, individual identities of chimpanzees from which sequences were generated are not known. In any given sampling location, samples yielding the same haplotype may therefore have come from the same individual. Accidental oversampling of individuals could effect the shape of the mismatch distribution. To test the effects of double sampling on the shape of the mismatch distribution, a reduced data set was created. Haplotypes within a sampling location which were identical were assumed to be so because of double sampling. All such identical haplotypes were collapsed. The size of the reduced data set under this "worst-case scenario" of 100% double sampling was 160 sequences. Figure 4.4 shows the mismatch distribution for the reduced data set, overlaid on that for the full data set. The shapes of the two distributions are virtually identical, and the values of τ and θ_0 are very similar (5.37 and 2.27 respectively in the full data set, and

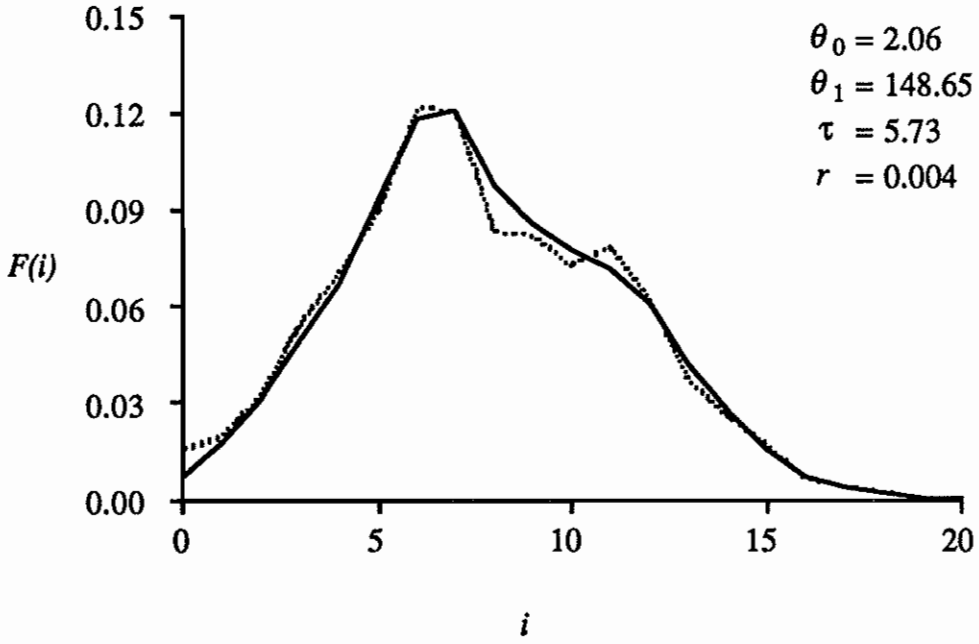


Figure 4.4: Mismatch distribution for reduced data set of 160 sequences (solid line) created by collapsing redundant haplotypes within sampling locations. Full data set (262 sequences; broken line) is superimposed for comparison.

5.73 and 2.06 respectively in the reduced data set). The value of θ_1 in the reduced data set (62.55) is considerably lower than that in the full data set (148.65). This is to be expected, since elimination of identical haplotypes in the reduced data set will necessarily decrease the 0th category of the distribution and thus lower its vertical intercept. Given the high concordance between mismatch distributions generated from these two data sets, and given the low demographic probability of double sampling (see discussion in Chapter 2), all subsequent analyses are presented for the full data set. Analyses using reduced data sets did not yield statistically different results from the full data set in any case.

Confidence regions

Figure 4.5 shows a confidence region built for the distribution shown in Figure 4.3. The confidence region is illustrated as a series of graphs, each representing a different degree of population growth, in which filled dots represent accepted values of θ_0 and τ. The population history used to generate this interval was one of panmixia and sudden expansion, in which no subdivisions were recognized and no migrants exchanged. This is the simplest population history that could be constructed for these data under a model incorporating population growth. The history takes the following form, which expresses population growth in time periods called "epochs," each with a duration of tau units of time (Rogers, 1995a; Rogers and Jorde, 1995):

	Theta	*Migration*	*tau*	*Subpopulations*
Epoch 2	θ_1	0	τ	1
Epoch 1	θ_0	0	Infinite	1

An infinite τ for Epoch 1 means that the population was demographically stable for a long time prior to expansion. The confidence interval suggests that the data are consistent with all degrees of population growth (levels of growth as high as 10^9 were tested, but are not

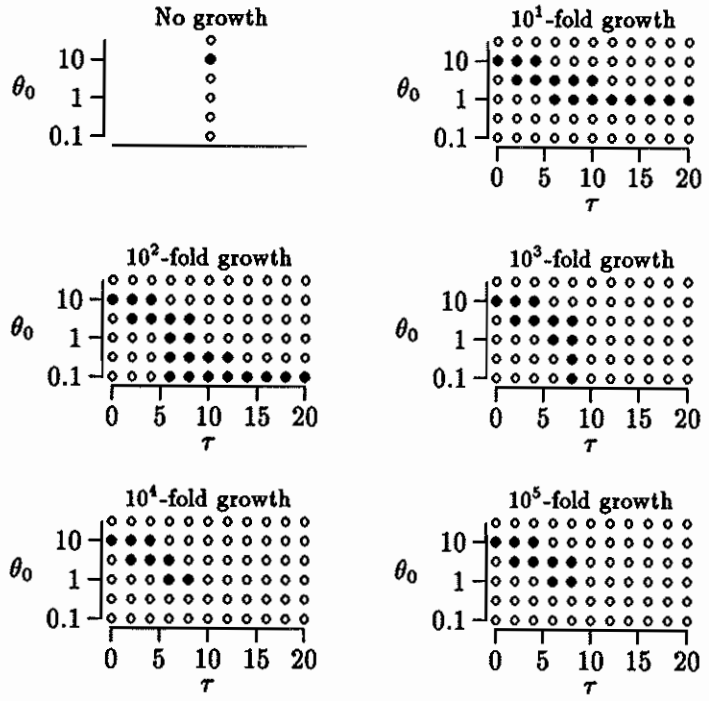


Figure 4.5: A 95% confidence region plotted for the data shown in Figure 4.3 of 262 eastern chimpanzee d-loop subregion sequences. Confidence intervals were calculated using 1000 simulated data sets under a population model assuming sudden expansion from size θ_0 to size θ_1 at time τ. The model also assumes that population subdivisioning and migration were absent (see text for full explanation). Filled circles represent points within the 95% confidence region; open circles represent points outside the region.

plotted), and with a broad range of values of θ_0 and τ. The inability of this technique to rule out scenarios of no or low population growth is particularly surprising considering the wave-like form of the distribution and the low value of *r*. By analogy to the human populations described by Sherry *et. al.* (1993), an *r* value as low as the one observed in this study would be consistent with at least 10^2 fold expansion. Ranges of θ_0 and τ are, however, narrower for levels of population growth greater that 10^2 and remain constant for levels of growth greater than or equal to 10^3. The confidence region also suggests an upper limit of 10 for θ_0, since no values greater than 10 appear for any degree of population growth tested.

The history of the East African chimpanzee population is not, however, likely to have been one of simple expansion and panmixia. Rather, as argued in Chapter 1, the east African chimpanzee population has likely experienced repeated, cyclic expansions and contractions corresponding to climatically-induced changes in forest cover. Such a scenario would result in one or several bottleneck events, each coinciding with maximum glaciation and minimal forest cover. Figure 4.6 shows a confidence region constructed under the simplest such model of population history, a single bottleneck. This model takes the following form (both values of τ are identical):

	Theta	*Migration*	*tau*	*Subpopulations*
Epoch 3	θ_1	0	τ	1
Epoch 2	θ_0	0	τ	1
Epoch 1	θ_1	0	Infinite	1

Here it is assumed that a bottleneck occurs at time 2τ and lasts for τ units of time, and that the population size during the bottleneck is θ_0. The confidence region generated under this population history also encompass a wide range of values of θ_0 and τ, and is generally comparable to that in Figure 4.5. However, no accepted parameter values appear in the "No-growth" section of the confidence region, implying that some degree of

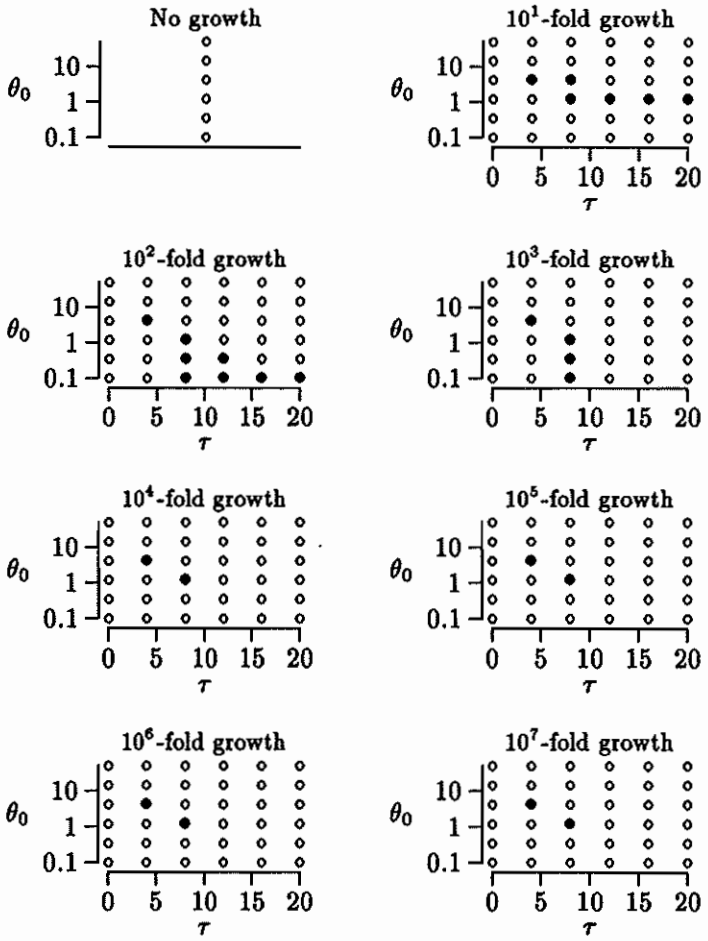


Figure 4.6: A 95% confidence region plotted for the data shown in Figure 4.3 under a population model assuming a bottleneck of duration τ, followed by expansion at time τ. Subdivisioning and migration were assumed to be absent. Confidence intervals were calculated using 1000 simulated data sets.

population expansion is supported by the data under the model of growth just described. Again, an upper limit of approximately 10 is defined for θ_0.

Contractions and expansions in forest cover during the Pleistocene have probably also resulted in the repeated vicariance of chimpanzee habitat, leading to subdivisioning within the chimpanzee population. It is therefore unrealistic to assume, as the confidence regions described above do, that the east African chimpanzee population mates randomly. More likely, the post-expansion population has consisted historically of a series of linked subpopulations which exchange migrants in proportion to their geographic proximity and the contiguity of forest between them (see Chapter 3). Figure 4.7 plots a confidence region under an extreme form of population subdivisioning. The population history used to generate this confidence region is the following:

	Theta	*Migration*	*tau*	*Subpopulations*
Epoch 2	θ_1	0	τ	13
Epoch 1	θ_0	0	Infinite	1

This model postulates that chimpanzees have expanded out of a single Pleistocene refuge into thirteen post-expansion subpopulations, representing the eight eastern forests included in the analysis plus the five Zaïrian locations. The lack of migration among subpopulations is unrealistic, but delimits one extreme of a continuum of degrees of population subdivisioning. The confidence interval generated under this model does not differ significantly from those in Figures 4.5 and 4.6. All levels of population expansion are accepted, as are a broad range of values of θ_0 and τ. Again, the confidence region delimits an upper bound of 10 for θ_0. Confidence regions were also calculated for three levels of inter-subpopulation migration (0.01, 0.1, and 1 migrants per generation), representing low, medium and high migration, respectively. The resulting confidence regions (not shown) do not differ from the regions presented in Figures 4.5 - 4.7. Ranges

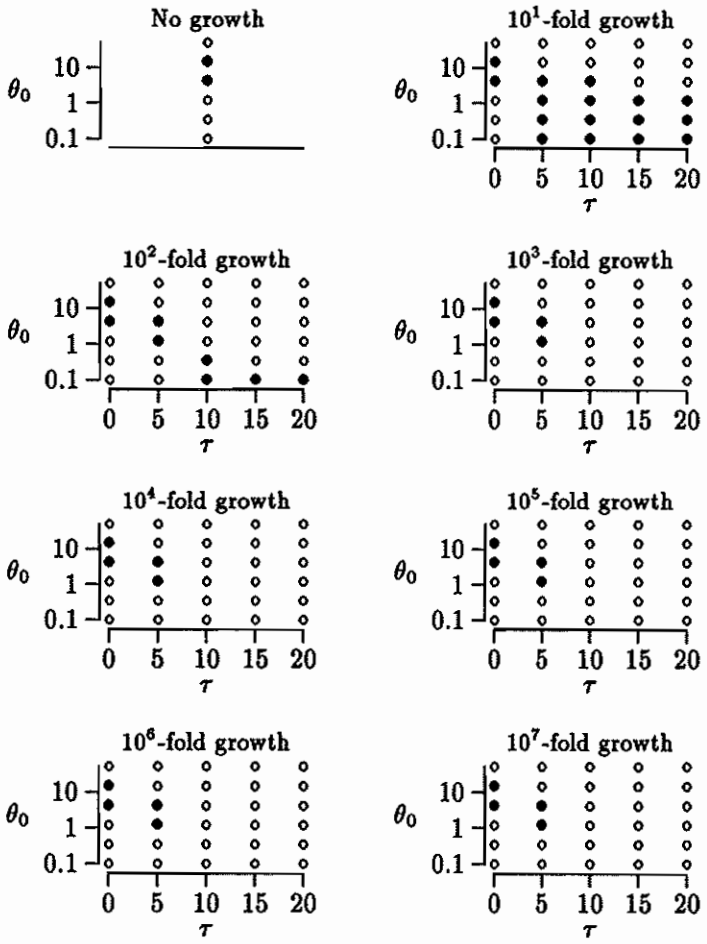


Figure 4.7: A 95% confidence region plotted for the data shown in Figure 4.3 under a population model of rapid expansion at time τ from a single population into a post-expansion population consisting of 13 isolated subdivisions. Migration among subdivisions was assumed to be absent. Population histories incorporating greater degrees of migration produced confidence regions similar to the one shown. Confidence intervals were calculated using 1000 simulated data sets.

of θ_0 and τ are wide, and no values of θ_0 are greater than approximately 10.

Although the population histories modeled above do not serve to test the full range of hypotheses about chimpanzee paleodemography, they show that more refined modeling is not likely to alter the general results obtained from confidence region analysis. Confidence regions do not in general differ significantly even across a broad range of population histories, representing extreme levels of expansion, subdivisioning and migration (Rogers, 1995c). More intricate population histories than those tested are likely given the complexity of forest expansions and contractions in East Africa during the Pleistocene. However, the range of plausible population histories under such models is too large to be analyzed thoroughly. Patterns of population expansion more complex than sudden growth and panmixia may be consistent with the data, but should affect neither the general shape of the mismatch wave nor the accuracy of parameter values obtained from confidence region analysis. This property holds even for confidence regions generated under models of population growth incorporating simplifying assumptions about subdivisioning and migration (Rogers, 1995c).

Geographic patterning in mismatch distributions

If the eastern chimpanzee population is free of significant population structure, then samples of individuals from different locations should, within the bounds of sampling error, be representative of the population as a whole. In other words, a mismatch distribution generated using individuals from any one sampling locality should retain the general properties of the mismatch distribution in Figure 4.3, assuming that the geographic origin of a sequence has no bearing on its phylogenetic positioning within the eastern chimpanzee gene tree.

To test this hypothesis, mismatch distributions were generated for each of the sampling localities listed in table 2.1, and depicted in Figure 2.1. Repeated haplotypes

within sampling locations were retained based on the results shown in Figure 4.4. Collapsing identical haplotypes within sampling locations did not significantly alter the results, which are presented in table 4.1. Table 4.1 includes a mismatch distribution for data from Gombe (GEKA, eliminated from the previous analyses) for the sake of comparison. GEKA sequences were aligned by hand with reference to sequences generated in this study, and all positions containing missing data were omitted. The mismatch distribution for GEKA is thus not entirely comparable to those from other localities, in that it is based on a subset of the nucleotide positions available for other localities (219 bp). A mismatch distribution is also described for the Entebbe Zoo population (EEZO), even though the geographic origins of these animals are unknown.

The parameters defining the shapes of the individual mismatch distributions vary considerably among sampling locations. θ_0 varies from 0 to 5.5, with a mean of 2.37 and a standard deviation of 1.46. θ_1 varies from 1.79 to 80.25 with a mean of 14.97 and a standard deviation of 18.92, and τ varies from 0.31 to 7.27, with a mean of 3.98 and a standard deviation of 1.80. Values of r span a range consistent with population expansion in some cases and long-term population stability in others (0.01 to 0.405; mean = 0.117, standard deviation 0.1; undefined values, shown as ">1.0" in Table 4.1, excluded). 95% confidence intervals are also presented for values of τ. Confidence intervals were constructed from 1000 iterated data sets according to the method of Rogers (1995a) previously described, under the assumption of 10^2 fold population expansion. The population history used was one of simple expansion from size θ_0 to size θ_1 at time τ. Parameter values used for each population were empirical values for the mismatch distribution from that population. Confidence intervals built for other degrees of population expansion did not differ appreciably from those reported for 10^2 fold growth. More complex population histories were not investigated, for the reasons outlined above.

The broad range of parameter values reported in Table 4.1 implies that further

Table 4.1: Parameters of mismatch distributions for individual sampling locations

	*Location**	*n*	θ_0	θ_1	τ	95% C.I. (τ)†	*r*
1	BAMA	13	2.58	1.79	0.31	$0.00 < \tau < 2.30$	0.121
2	BOPI	12	1.63	10.00	3.79	$1.25 < \tau < 7.38$	0.072
3	BOSO	15	0.88	7.75	4.45	$4.13 < \tau < 6.75$	0.098
4	EEZO	26	2.44	80.25	4.76	$1.25 < \tau < 9.25$	0.010
5	GEKA	19	2.15	13.25	4.89	$3.25 < \tau < 8.80$	0.028
6	IARA	12	2.02	10.00	3.20	$0.75 < \tau < 6.63$	0.108
7	IIAA	13	2.80	18.50	7.27	$7.00 < \tau < 11.13$	0.060
8	IIAE	12	2.76	2.00	1.33	$0.13 < \tau < 2.54$	0.405
9	IIAW	13	2.08	12.00	5.57	$4.88 < \tau < 8.75$	0.050
10	IILA	17	1.18	16.00	5.30	$3.88 < \tau < 8.25$	0.043
11	KEDN	14	3.89	2.64	2.86	$0.13 < \tau < 5.60$	0.243
12	KEKA	15	0.00	7.75	3.33	$2.25 < \tau < 6.25$	> 1.0
13	KEKU	14	3.41	4.69	4.16	$0.63 < \tau < 7.88$	0.112
14	KENO	13	4.95	4.20	2.05	$0.00 < \tau < 3.63$	0.205
15	KUSL	13	2.19	38.00	3.20	$0.00 < \tau < 7.63$	> 1.0
16	NESN	13	0.28	10.14	6.08	$6.00 < \tau < 7.20$	0.222
17	RIKA	13	3.77	38.00	5.39	$0.00 < \tau < 10.75$	0.032
18	SIMU	13	5.49	2.25	1.25	$0.00 < \tau < 2.00$	> 1.0
19	SINI	11	2.49	10.00	5.11	$3.13 < \tau < 8.88$	0.083
20	TOBA	10	0.46	10.25	5.63	$3.88 < \tau < 9.75$	0.097
21	*Total*	262	2.27	62.55	5.37	$2.63 < \tau < 7.00$	0.005

* For full description of sampling locations, see Table 2.1 and associated text.
† 95% confidence intervals around tau; see text for full description.

investigation of location-specific mismatch distributions might prove informative. Biogeographic evidence suggests that lowland rainforests in Zaïre and their indigenous fauna share different histories than peripheral insular forests on the eastern borders of the chimpanzee's range (see discussion in Chapter 1). Mismatch distributions for sequences from Zaïrian forests (n = 65) and for sequences from eastern forests (n = 171) were therefore calculated separately, and are shown in Figure 4.8. One striking feature of these distributions is the disappearance of bimodality in the Zaïrian distribution and its enhancement in the eastern distribution. The eastern mismatch distribution shows a marked second mode at $i = 11$, which corresponds to that in the full-data mismatch distribution shown in Figure 4.3. The comparative smoothness of the Zaïrian distribution is reflected in its slightly lower r value (0.006, compared to 0.007 for eastern distribution), although both values are consistent with large degrees of population expansion. The values of τ for the two distributions also differ (4.82 for the eastern distribution versus 6.16 for the Zaïrian distribution). A higher value of τ for the Zaïrian distribution implies that expansion occurred earlier in Zaïre than in the east, as would be expected if chimpanzees from lowland Zaïrian forest refugia dispersed eastward during post-glacial forest expansion (see Chapter 1). However, 95% confidence intervals surrounding these estimates show considerable overlap ($4.0 < \tau < 9.5$ for the Zaïrian distribution; $1.75 < \tau < 8.5$ for the eastern distribution; see Table 4.2).

The observation that bimodality is localized to the eastern forests raises the question of whether it is further localized. Mismatch distributions were therefore constructed for each of the nine eastern forests for which data were available. When a forest was represented by more than one sampling location (Budongo, Kibale and Semliki), data from each sampling location within the forest were combined. Gombe data were included to allow qualitative comparisons to be made with other populations. Results are shown in Figure 4.9. Mismatch distributions were defined as bimodal if they contained a peak at any value of $i \geq 11$ which was equal to or greater than the largest peak in the distribution for

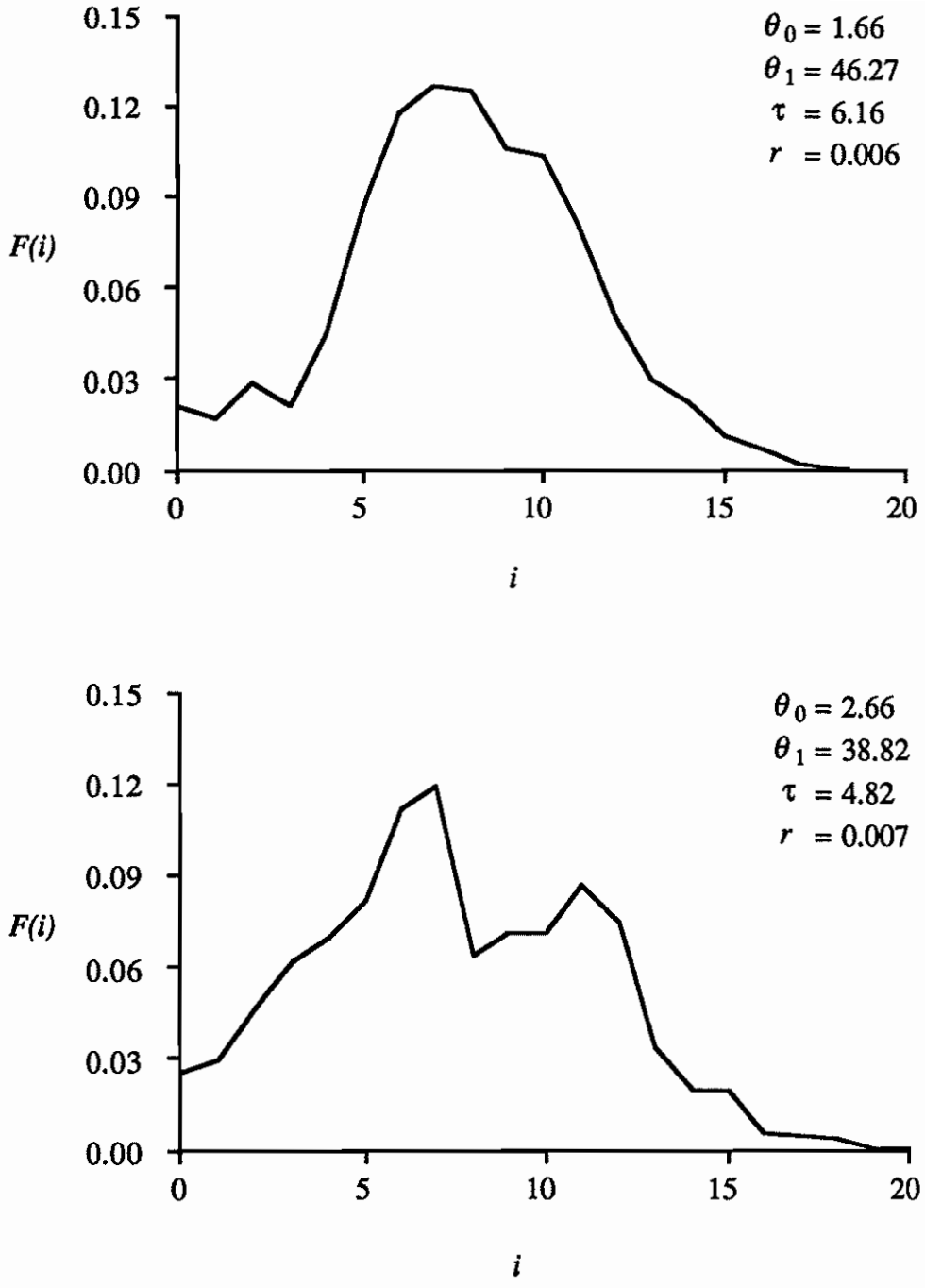


Figure 4.8: Mismatch distributions for Zaïrian forests (upper panel; n = 65 sequences) and eastern forests (lower panel; n = 171 sequences).

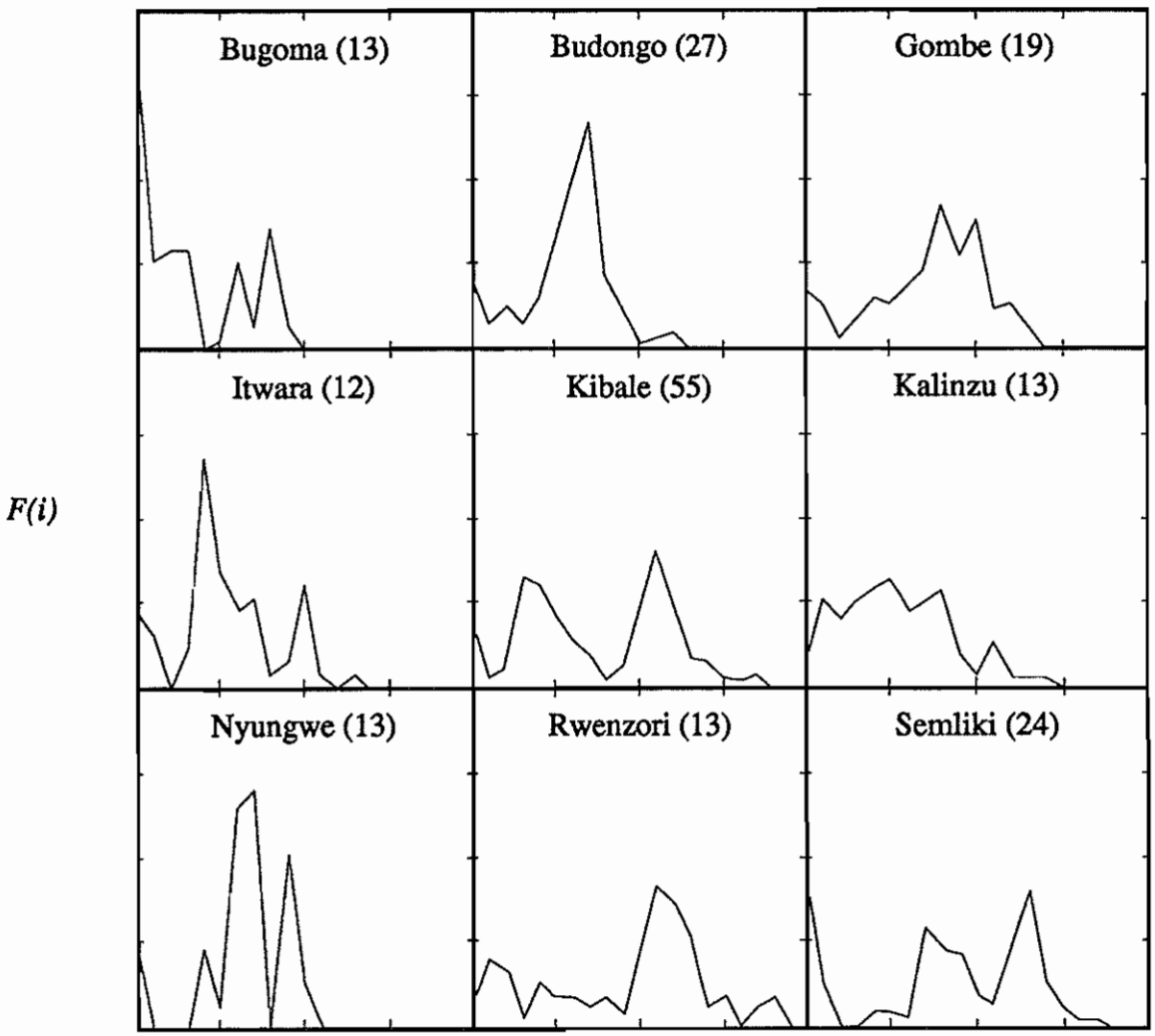


Figure 4.9: Mismatch distributions for nine eastern forests, plotted on the same scale (vertical axes: $F(i) = 0$ to 0.4; horizontal axes: $i = 0$ to 20). Sequences from separate locations within forests were combined when such data were available (for Budongo, Kibale and Semliki). Numbers in parentheses are sample sizes.

lower i values. The criterion of a peak at $i \geq 11$ was chosen to reflect the observation that the full and eastern forest data sets contained a second mode at this value of i.

These criteria clearly differentiate two sets of mismatch distributions. Distributions from Kibale, Rwenzori and Semliki are bimodal; distributions from Bugoma, Budongo, Gombe, Itwara, Kalinzu and Nyungwe are unimodal (by the above definition). In Kibale and Rwenzori, the higher mode corresponds exactly to $i = 11$. In Semliki, the higher mode corresponds to $i = 13$. In all three cases, the higher mode is also the principal mode in the distribution. These results are particularly striking in view of the large combined sample size for Kibale (55), and the moderately large sample size for Semliki (24). Bimodality in these distributions does not likely result from sampling error.

The prominent mode at $i = 11$ for the eastern forest distribution (Figure 4.8, lower panel) is not, however, a simple additive result of the individual modes from the Kibale, Semliki and Rwenzori distributions. Rather, it results also from inter-forest interactions not evident in the separate distributions plotted in Figure 4.9. Data were therefore recombined for all eastern forests showing unimodal distributions, and for the three forests showing bimodal distributions. Results are shown in Figure 4.10. The recombined data set for eastern unimodal forests retains its unimodality, while the recombined data set for eastern bimodal forests is bimodal by the aforementioned criterion. In the eastern unimodal distribution, there is no evidence whatsoever of a second mode at $i \geq 11$. By contrast, the eastern bimodal distribution contains its principal mode at $i = 11$.

Bimodality in the eastern forest data set, and therefore in the full data set, is localized to Kibale, Rwenzori and Semliki Forests. The combined data sets for eastern unimodal and eastern bimodal forests contain adequate sample sizes to rule out sampling error as a an explanation for the different shapes of their corresponding mismatch distributions. Values of θ_0, θ_1 and τ may be considered relatively accurate for the unimodal distribution. However, values of θ_0, θ_1 and τ should be viewed cautiously for

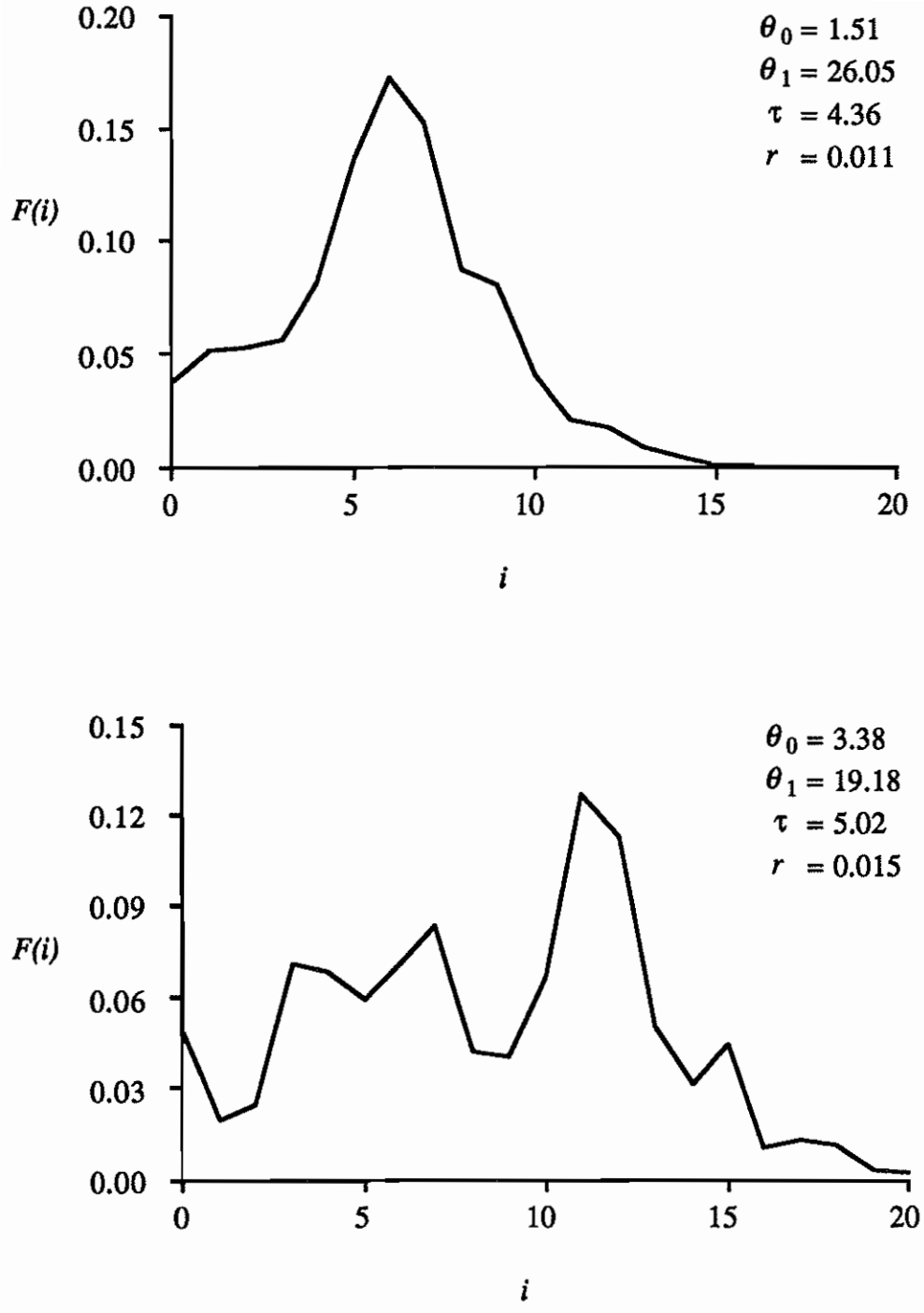


Figure 4.10: Mismatch distributions for combined data sets from eastern unimodal forests (upper panel; n = 78 sequences) and eastern bimodal forests (lower panel; n = 93 sequences).

the bimodal eastern distribution. Because the theoretical model underlying the calculations of these parameters assumes unimodality, theoretically-fit parameter values may not accurately reflect the shape of a markedly bimodal distribution. In particular, the mismatch distribution for bimodal eastern forests has its principal peak at $i = 11$, but a theoretical peak at $\tau = 5.0$ when fit to a unimodal distribution. Values of r for both distributions are nevertheless small enough to suggest histories of population expansion in both cases. It is also interesting to note that the eastern bimodal distribution is substantially broader than the unimodal distribution. No values of i exist in the eastern unimodal distribution which are greater than 16, whereas values up to $i = 20$ exist for the eastern bimodal distribution.

The analysis presented above demonstrates that haplotypes are not distributed across the landscape randomly with respect to their genealogical relationship to other haplotypes. Rather, three forests, Kibale, Rwenzori and Semliki, contain a substantially higher proportion of divergent chimpanzee haplotypes than do other locations. This difference is indicated by the presence of a second, higher mode in the mismatch distributions within these individual forests, which is retained in the combined mismatch distribution between them. This difference is also indicated by the higher maximum value of i for the distribution from these forests. This trend is not evident in the phylogenetic trees presented in Chapter 5, again indicating the complementarity of mismatch distribution analysis and tree-based analyses.

Biogeographic implications of bimodality in mismatch distributions

The results described for chimpanzee mismatch distributions are particularly intriguing since the three bimodal eastern forests (Kibale, Rwenzori and Semliki) are geographically clustered (see Table 2.1 and Figure 2.1). Rwenzori and Semliki Forests are presently contiguous, and Kibale Forest was most likely also contiguous within the last few thousand years (Hamilton, 1984). It is therefore probable that similarities among these

three forests reflect common history, rather than coincidence. This hypothesis is further strengthened by the retention of bimodality in the combined mismatch distribution for these three forests, and by the fact that values of *i* greater than 16 occur only in these forests (see Figure 4.9). In short, Kibale, Rwenzori and Semliki contain markedly higher proportions of divergent chimpanzee haplotypes than do other forests.

Marjoram and Donnelly (1994) argue that bimodal mismatch distributions occur when the phylogeny of the DNA sequences analyzed contains two (or more) ancient lineages, significantly predating the population expansion event. The statistical probability with which such lineages persist over time is a function of the degree to which the population is subdivided, and the size of the population prior to (and after) expansion. A population which is subdivided and experiences low levels of migration between subdivisions is likely to generate a multimodal mismatch distribution. Subdivisioning increases the likelihood that different ancient lineages will be preserved in each subdivision. Pairwise comparisons between these ancient lineages will generate the higher mode of a bimodal mismatch distribution. Sequence differences within each lineage will generate the lower mode. This pattern is dependent on limited migration between subpopulations. High levels of migration lower the probability with which separate ancient lineages are preserved and encourage unimodality even in subdivided populations.

Bimodal mismatch distribution may, however, arise even in panmictic populations. Two or more ancient lineages may be retained even in a nonstructured population, if the size of the pre-expansion population is large enough to offset the probability of ancient lineages becoming extinct. This scenario would occur in a population experiencing a long period of constant, but substantial, population size, followed by rapid expansion. The amount of time during which the population is of constant size, relative to the recency of the expansion event, determines the probability with which multiple ancient lineages are retained. A relatively large pre-expansion population, coupled with very recent population

expansion, will thus favor bimodality. Bottleneck effects cause much the same result in that they mimic the effects of pre-expansion population size constancy.

The bimodality of mismatch distributions from Kibale, Rwenzori and Semliki indicate the localized preservation of ancient chimpanzee lineages. Because of the geographic contiguity of these forests (Howard, 1991), population subdivisioning among them is unlikely to account for this pattern. This conclusion is also supported by the analyses in Chapter 3, which demonstrate a conspicuous lack of subdivisioning even over much greater geographic distances. Semliki, Rwenzori and Kibale therefore likely share a history in which the pre-expansion population was sufficiently large to enhance the probability that any chimpanzee lineages, including ancient ones, were retained. This would be the case if the forests were near to a Pleistocene refuge. Under this scenario, these three forests would have been the first to come into contact with the refuge population during post-glacial forest expansion. Proximity to the refuge would thus have facilitated immigration, and the introduction into the populations of a representative subset of "refuge" haplotypes. If divergent haplotypes were also rare, they would have dispersed into Semliki, Rwenzori and Kibale with higher probability than into other, more distant populations. Peripheral locations would, in other words, have lost these rare haplotypes due to the cumulative effects of a succession of founder events (Mayr, 1954; Carson, 1983). The current rarity of divergent haplotypes is evidenced by the fact that, in the full data set mismatch distribution (Figure 4.3), values of $i \geq 11$ comprise only about 15% of the area under the mismatch curve. It is also evident in the phylogenetic trees presented in Chapter 5, in which the most divergent clade of eastern chimpanzees is represented by only 3 to 11 haplotypes (2-9% of all haplotypes). Such a scenario would explain both the unimodality of the peripheral eastern forest mismatch distribution and the overall signature of sudden population expansion evident in all subsets of the data. The founder effect hypothesis could also explain the lack of bimodality in Itwara Forest, which, although geographically clustered with Kibale, Semliki and Rwenzori, is small and relatively

species-poor (Howard, 1991).

Independent biogeographic data suggest that Kibale, Semliki and Rwenzori differ from the other forests examined in this study in having unusually diverse mammalian and primate communities. Kibale Forest is known for its diversity of anthropoid primates (Struhsaker, 1981; Howard, 1991). Kibale is also the only Ugandan forest containing the red colobus subspecies *Colobus badius tephrosceles* (Struhsaker, 1975). A related subspecies of red colobus, *C. b. ellioti*, is found in Uganda only in the Semliki Forest (Haddow, 1952; Colyn, 1991; and personal observation). Semliki, like Kibale, also represents one of the richest forests in Africa for primates (Struhsaker, 1981). Semliki differs from Kibale in containing species endemic to lowland Zaïrian rainforest (notably *Cercopithecus neglectus* and *Cercopithecus pogonias* for the primates), which demonstrate Semliki's zoogeographic affiliations with the eastern Zaïrian forest block (Howard, 1991). Little is known about the fauna of the medium-altitude forests surrounding the Rwenzori mountains (Howard, 1991). Noteworthy in the context of the present study is the existence of the endemic black and white colobus *Colobus angolensis ruwenzorii*, found only in the Rwenzori mountains (Struhsaker, 1981). The Rwenzori are also noted for their unique montane vegetation, unusually high rainfall, and a remarkably diverse mammalian community, second in East Africa only to Semliki Forest (Osmaston, 1967; Rodgers, Owen and Homewood, 1982; Howard, 1991).

Hamilton (1976) has proposed two major Pleistocene refugia in the vicinity of Uganda on the basis of such data: one near Bwindi Forest in the southwest of the country, and the other immediately to the west of the Rwenzori Mountains. Struhsaker (1981) and Rodgers *et. al.* (1982) provide support for this hypothesis by demonstrating that species numbers of primates and mammals, respectively, decline in East African forests with distance from the Rwenzori Mountains. Colyn (1987, 1991) shows that a Rwenzori-localized montane refuge has, as part of a linked system of afromontane altitudinal refuges,

served as a major center of speciation and origin of westward dispersal for primates. Kingdon (1981) has recognized the importance of the Rwenzori as a stepping stone for the eastward dispersal of African mammals. A refuge near the Rwenzori could have resulted from the role of the mountains as a local watershed (Struhsaker, 1981). If current patterns of relatively high rainfall in the Rwenzori have persisted throughout colder, more arid glacial periods, then the Rwenzori themselves may have fostered locally wet conditions. Such conditions would have favored the localized persistence of chimpanzee foods such as terrestrial herbs, which have been an evolutionarily-important component of the chimpanzee diet (Malenky and Wrangham, 1994) and occur in very high abundance in the Rwenzori today (e.g. *Afromomum mala*; pers. obs.).

Data from the present study are, however, inconsistent with the notion that a major Pleistocene refuge for eastern chimpanzees existed in northeastern Zaïre (Grubb, 1982; Colyn, 1991). Colyn (1991) has reconstructed a series of fluvial refuges which, in conjunction with the montane refugia of western Rift, contribute to mammalian species-richness in northeastern Zaïre. Under this model, the chimpanzee population inhabiting Zaïrian lowland forests during glacial periods should have been large. The high incidence of speciation inferred under such conditions implies that the Zaïrian forest was geographically subdivided, perhaps by fluvial barriers (Colyn, 1991; Ayres and Clutton-Brock, 1992). This, in turn, implies that the chimpanzee population in Zaïre during such periods may have been subdivided. A large and subdivided pre-expansion population would predict a bimodal mismatch distribution for Zaïrian forests. The conspicuous absence of bimodality in the Zaïrian distribution is difficult to explain given this prediction. However, the larger value of τ for the Zaïrian distribution (6.16) than for the eastern unimodal distribution (4.36) is consistent with the notion that Zaïrian populations are older than peripheral eastern forests. The largest value of τ for any single population in the study (7.27) was observed for a population from the Ituri Forest, Zaïre (IIAA), an acknowledged

center of endemism and species diversity which is traditionally reconstructed as an important Pleistocene refuge (Grubb, 1982). However, within Zaïre, the maximally divergent haplotypes differ at only 18 sites, which is less than the maximum within-population value observed in the study ($i = 20$ for RIKA population; see Figure 4.9). This observation, plus the lack of bimodality in the Zaïrian mismatch distribution, makes inferences about the locations of refugia problematic.

Population sizes and dating of expansion events

Estimating population sizes and dates of population expansion on the basis of mismatch distributions is fraught with difficulty. Errors surrounding estimates of θ_0, θ_1 and τ are large, and reduce the precision of any estimates generated. Equally problematic is the lack of accuracy with which mutation rates are known. Nucleotide divergence rate estimates vary widely, and standard errors around most estimates are unknown (Vigilant *et. al.*, 1991; Ward *et. al.*, 1991). This is especially true for the hypervariable d-loop regions, where multiple substitutions have confounded accurate rate estimates (Ruvolo *et. al.*, 1993). Until these shortcomings are resolved, dates and population sizes derived from mismatch distribution data should be considered rough estimates only.

For the present analysis, estimates of nucleotide divergence rates will be taken from Vigilant *et. al.* (1991) and Ward *et. al.* (1991). Both studies derive nucleotide divergence rates using DNA sequence data. The sequences used are from a segment of the human mitochondrial control region homologous to that analyzed in the present study for chimpanzees. Furthermore, both studies compare chimpanzee sequences to human sequences to estimate nucleotide divergence rates within humans. Assuming evolutionary rate constancy between humans and chimpanzees (Kocher and Wilson, 1991), these estimates should therefore be equally reasonable for within-chimpanzee evolution. Finally, estimates from Vigilant *et. al.* (1991) and Ward *et. al.* (1991) have been used in previous studies of modern human evolution based on mismatch distribution analysis (Rogers and

Harpending, 1992; Sherry *et. al.*, 1994; Rogers and Jorde, 1995). These estimates are therefore useful from a comparative perspective. The estimate from Vigilant *et. al.* (1991) of a nucleotide divergence rate of 11.5% per million years is based on a transition: transversion ratio of 15:1, and a date of divergence between humans and chimpanzees of 6 million years. It will be used as a low-end estimate of the actual nucleotide divergence rate. The estimate of Ward *et. al.* (1991) of 33% per million years is based on a transition:transversion ratio of 30:1 and a date of divergence between humans and chimpanzees of 4 million years (almost certainly too recent). This estimate will be used as a high-end estimate of the actual nucleotide divergence rate.

Dates and population sizes were calculated following the method outlined in Rogers and Jorde (1995). The following example illustrates the method, using Vigilant's rate estimate of 11.5% per site per million years and the value of $\tau = 5.37$ for the eastern chimpanzee mismatch distribution. Vigilant's rate estimate is the *between lineage* rate, and is divided by two to yield a *within lineage* rate in equation 4.5. This rate must also be multiplied by the generation time (25 years per generation) to yield a *per generation* rate of 2.9×10^{-6} nucleotide substitutions per site per generation. From equation 4.4:

> $u = 368 \times 2.9 \times 10^{-6} = 1.1 \times 10^{-3}$ substitutions/site/generation for the 368 base pair length region studied here.

From equation 4.5 (assuming $\tau = 5.37$):

$$t = 5.37/(2 \times 1.1 \times 10^{-3}) = 2441 \text{ generations}$$

and from equation 4.6:

$$T_{(\text{abs. yrs.})} = 2441 \text{ generations} \times 25 \text{ years per generation} = 61{,}023 \text{ years}$$

All date and population size estimates were calculated in this fashion.

Confidence region analysis implies that θ_0 has an upper bound of approximately 10 under all models of population growth tested. Assuming a nucleotide divergence rate of 11.5% (as in the example above), a conservative upper-limit estimate of the pre-expansion population size is roughly 6000 breeding females. A "median" estimate, based on an intermediate nucleotide divergence rate of 22.25% and the empirically-derived value of 2.27 for θ_0 in the full data set, implies a pre-expansion population size of approximately 700 breeding females. If, as suggested by the low value of the roughness statistic (r = 0.005 for the full data set), the chimpanzee population has experienced at least 100-fold population growth, then the size of the post-expansion population would be approximately 70,000, with an upper limit of 600,000 breeding females. In this context, it is interesting to note that an independent estimate of the present number of breeding females in *P. t. schweinfurthii* of 22,000 (Teleki, 1989) is lower than the value reported here. Teleki's estimate is based on the current area of potential chimpanzee habitat within the subspecies range. The estimates presented in this study, however, reflect population numbers just after population expansion, probably at approximately 10 kya (see Chapter 1 and below). The discrepancy between Teleki's estimate and the estimate presented here must partly reflect recent changes in chimpanzee population numbers. Such changes may be due to human-induced habitat loss, which has been extensive since the first occurrence of iron tools in the East African archaeological record at approximately 2.5 kya (Phillipson, 1977; Hamilton, 1984; Hamilton, Taylor and Vogel, 1986). However, since no lower bound was delimited for θ_0 based on confidence region analysis, Teleki's estimate may be consistent with the data presented in this study.

Rough confidence limits around dates of population expansion were estimated using 95% confidence regions around τ (calculated as described above) in conjunction with low- and high-end estimates of nucleotide divergence rates. Dates are calculated as described above, and are presented for various subpopulation of the full data set in Table

4.2. The full data set of 262 sequences ($\tau = 5.37$; see Figure 4.3) yields an estimate of the date of chimpanzee expansion in eastern Africa of 31.5 kya (using the median nucleotide divergence rate of 22.25%), with a rough confidence interval of between 10.8 and 82.7 kya using the low and high end rates described above in conjunction with 95% confidence regions. However, as discussed above, the geographic localization of bimodality suggests that different subpopulations of the data may have experienced different population histories. Therefore, separate consideration of mismatch distributions for individual forests may give a more informative picture of dates of population expansion.

Figure 4.11 presents "best guess" dates of population expansion, and approximate confidence intervals, for the nine eastern forests sampled (including Gombe), the five Zaïrian locations sampled, and the combined data set of 262 chimpanzee sequences. Table 4.2 presents numerical date estimates for subpopulations of the full data set based on a range of mutation rates and on location-specific confidence intervals around τ. Zaïrian populations appear slightly older than eastern populations ($\tau = 6.16$ for Zaïrian forests and 4.82 for eastern forests; see Table 4.2 for corresponding date estimates). The ranges of dates of expansion for Zaïrian and eastern populations overlap, however.

Dates based on unimodal mismatch distributions (Zaïre and unimodal eastern forests) are relatively reliable. However, dates based on bimodal distributions (e.g. the eastern bimodal forests) may be misleading. The values of τ on which these estimates are based (calculated from a mathematical model assuming unimodality) do not accurately reflect the shape of the empirical distribution, which is bimodal. The lower mode of this distribution corresponds approximately to the principal mode in the eastern unimodal distribution, and probably reflects the same population expansion event. The higher mode, however, is approximately 2.2 times the value of the first. A value of $\tau = 11$ would correspond to a date of expansion of 67.2 kya, using the median divergence rate of 22.25%. Using the highest rate (33%) gives an expansion date of 45.3 kya; using the

Table 4.2: Mismatch distribution parameters and dates of expansion for subpopulations of the full data set

	Population	*n**	θ_0	θ_1	τ	T1**	T2**	T3**	*r*	95% C.I. (τ)†	Date of expansion event (kya)‡
1	Kibale	56	3.53	11.73	3.76	42.76	14.90	22.10	0.038	$0.00 < \tau < 6.00$	$00.0 < t < 71.0$
2	Budongo	27	1.14	11.10	4.45	50.53	17.61	26.11	0.056	$4.00 < \tau < 6.80$	$16.5 < t < 80.0$
3	Semliki	24	4.13	4.52	3.82	43.42	15.13	22.44	0.055	$0.25 < \tau < 5.50$	$01.0 < t < 65.0$
4	Zaïre	65	1.66	46.27	6.16	69.98	24.39	36.17	0.006	$4.13 < \tau < 9.25$	$17.0 < t < 109.0$
5	Eastern Forests	171	2.66	38.82	4.82	54.81	19.10	28.33	0.007	$1.75 < \tau < 8.50$	$07.2 < t < 100.0$
6	Bimodal Eastern Forests	93	3.38	19.18	5.02	57.08	19.89	29.50	0.015	$2.75 < \tau < 7.00$	$11.3 < t < 83.0$
7	Unimodal Eastern Forests	78	1.51	26.05	4.36	49.57	17.27	25.62	0.012	$2.75 < \tau < 7.00$	$11.3 < t < 83.0$
8	All Data	262	2.27	62.55	5.37	61.02	21.25	31.52	0.005	$2.63 < \tau < 7.00$	$10.8 < t < 83.0$
9	*Homo sapiens****	135	1.29	214.40	8.34	94.80	33.03	48.98	0.004	$6.50 < \tau < 12.00$	$25.7 < t < 136.4$

* Sample sizes reflect combined data from all relevant sampling locations

** Times are in kya, and are based on empirical values of tau. T1 uses Vigilant *et. al.*'s (1991) nucleotide divergence rate of 11.5% per site per million years ($u = 0.0011$ mutations/site/generation); T2 uses Ward *et. al.*'s (1991) divergence rate of 33% per million years ($u = 0.0032$ mutations/site/generation); T3 uses the intermediate rate of 22.25% per million years ($u = 0.0021$ mutations/site/generation). See text for full explanation.

*** Human data taken from Vigilant et. al. (1991); see Figure 4.12.

† 95% confidence intervals around tau; see text for full description.

‡ Approximate confidence interval around dates of expansion calculated using 95% confidence limits around tau in conjunction with per site divergence rate estimates of 33% per million years and 11.5% per million years, respectively. The interval given here thus represents the low and high endpoints of the 95% confidence interval, dated using the high- and low-end divergence rates, respectively.

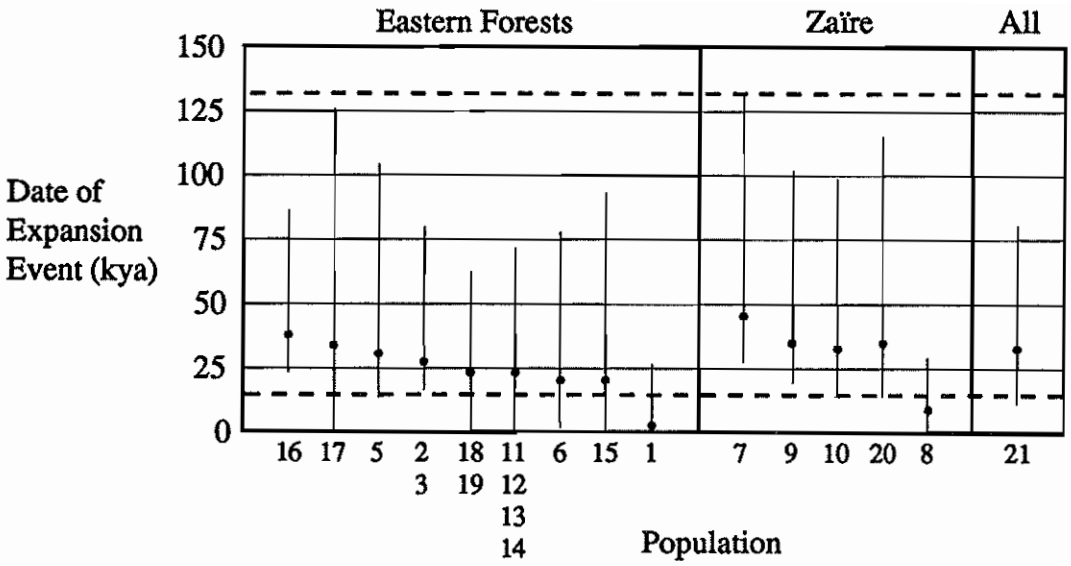


Figure 4.11: Dates of expansion for populations listed in Tables 4.1 and 4.2. Data from eastern forests represented by more than one sampling location (Budongo, Kibale, Semliki) were combined. Error lines represent confidence intervals based on extreme ends of 95% confidence limits around τ and nucleotide divergence rates of 11.5% (lower bound) and 33% (upper bound). Filled circles represent date estimates based on a mean nucleotide divergence rate of 22.25% and population-specific estimates of τ (see Tables 4.1 and 4.2). Population numbers refer to sampling locations listed in Table 4.1. Broken lines represent major deglaciation events.

lowest rate (11.5%) gives an expansion date of 130.0 kya for the lineages represented by this mode. To arrive at a time of origin for the eastern chimpanzee population, the above dating method was applied to the *maximum* pairwise difference observed within subpopulations (using the median divergence rate of 22.25%). This yielded an estimate of 117.5 kya for the origin of eastern forest populations (maximum $i = 20$), 111.6 kya for Zaïrian forest populations (maximum $i = 19$), and 117.5 kya for the full eastern chimpanzee sample (maximum $i = 20$). Using the low rate estimate of 11.5%, these times become 216 kya for the origin of Zaïrian populations, 227 kya for the origin of eastern populations, and 227 kya for the origin of the eastern chimpanzee population as a whole.

The ranges of date estimates given in Table 4.2 and Figure 4.11 are wide, but are generally consistent with a chimpanzee population expansion event corresponding to the last deglaciation beginning at approximately 12.5 kya (see Chapter 1). The dates given above are generally inconsistent with a population expansion corresponding to 130 kya or older, which would correspond to the previous deglaciation event. Two possible exceptions are the Rwenzori population (#17 in Figure 4.11) and the Ituri Forest, Afarama population (#7 in Figure 4.11). Potential dates of expansion for these populations extend notably farther back in time than do dates for other populations, and are marginally consistent with an expansion event at 130 kya. In this context, it is intriguing to note that both of these populations are located within areas corresponding to reconstructed Pleistocene refugia (Hamilton, 1976; Grubb, 1982).

If the population expansion event recorded for the majority of the other forests shown in Figure 4.11 does, in fact, correspond to an expansion at 12.5 kya, this implies that actual rates of nucleotide divergence may be much higher than previously reported (Vigilant *et. al.*, 1991; Ward *et. al.*, 1991). Assuming a date of expansion of 12.5 kya and values of τ between 2.63 and 7.0 (corresponding to the 95% confidence intervals calculated for the overall mismatch distribution), the actual rate of nucleotide divergence for

hypervariable region 1 of d-loop could be as low as 30% per nucleotide per million years, or as high as 79% per nucleotide per million years. Calculating dates from d-loop data is fraught with error specifically because of the high mutation rate of the region. As shown by Kocher and Wilson (1991) and Ruvolo *et. al.* (1993), multiple substitutions in this region necessitate that correction methods be used when deriving substitution rates. Unfortunately, the choice of correction methods can make a vast difference in the rates obtained, and thus in the dates inferred from them (Ruvolo *et. al.*, 1993).

Rogers and Jorde (1995) review the human genetic evidence and argue that the human population expansion occurred between 33 and 150 kya. They also argue specifically that eastern chimpanzees show evidence of coincident population expansion, and that this finding supports a causal link between the human and eastern chimpanzee expansion events. They base this conclusion on a mismatch distribution made from 37 *P. t. schweinfurthii* sequences, taken from Morin *et. al.* (1994a). Unfortunately, the sample size of chimpanzees was small, and the DNA sequences were largely incomplete. To test the validity of Rogers and Jorde's observation, mismatch distributions were created from the 123 distinct eastern chimpanzee haplotypes identified in the present study, and from 135 human haplotypes identified by Vigilant *et. al.* (1991). Results are presented in Figure 4.12. The two distributions are similar and show considerable statistical overlap when analyzed using the confidence-region techniques described above (see Table 4.2). The eastern chimpanzee distribution is, however, slightly skewed towards a higher τ value. Using the median rate estimate of 22.25% described above, this difference (1.36 τ units) corresponds to a time difference of 7.7 ky. Using the 11.5% rate estimate gives a difference in expansion times of 14.9 ky, while the highest rate estimate (33%) gives a difference of 5.2 ky.

Because confidence intervals around the date estimates given above overlap considerably, Rogers and Jorde's claim that a single climatic event drove both expansion events (they suggest the Toba volcanic eruption at 73.5 kya) cannot be rejected. However,

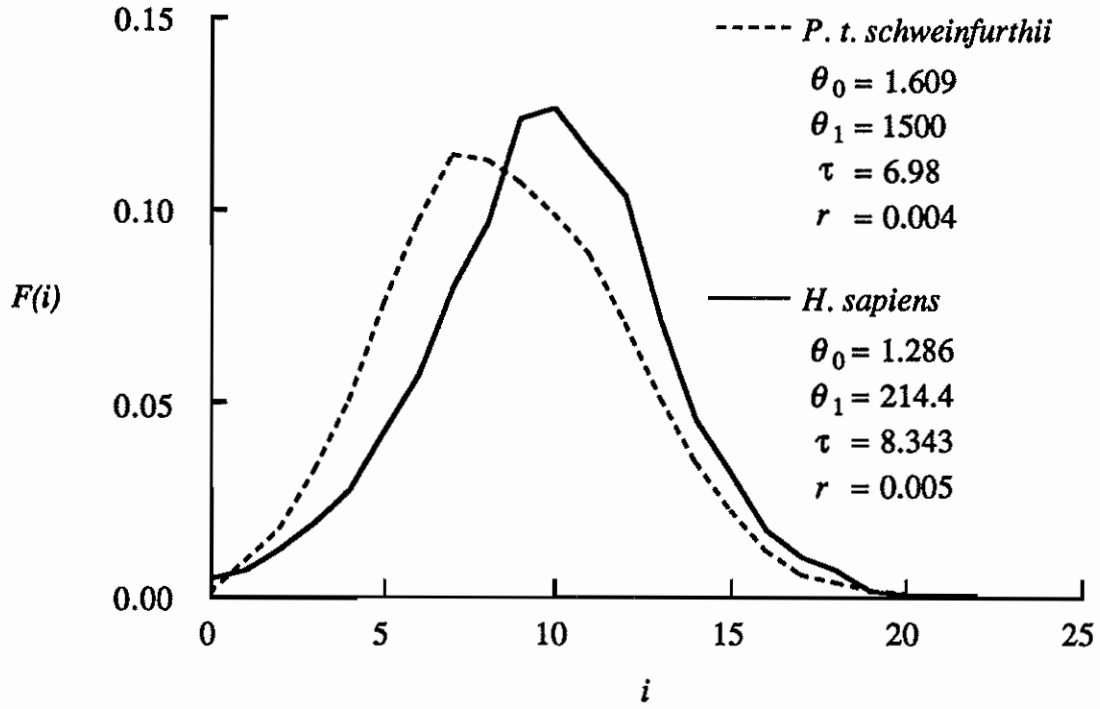


Figure 4.12: Mismatch distributions for humans and eastern chimpanzees. Human data consist of Vigilant's (1991) world sample of 135 mitochondrial control region sequences, edited to be homologous to the region sequenced in the present study. Chimpanzee data consist of 123 distinct haplotypes. All identical haplotypes within the chimpanzee sample were collapsed to make sample sizes approximately equal. The eastern chimpanzee mismatch distribution for the full data set is shown in Figure 4.3.

the broad range of statistically acceptable values of τ in both cases also demonstrates that such inference should be viewed with extreme caution. Once a population has undergone initial expansion, subsequent bottlenecks and expansions may not have any appreciable effect on the shape of the mismatch distribution (Rogers, 1995c). Therefore, any climatic event or events within the 95% confidence range of dates delimited by τ could be responsible for the wave-like form of the chimpanzee and human mismatch distributions. As Figure 1.3 shows, a great number of fine-scale climatic events have occurred within the confidence ranges of expansion dates delimited by the eastern chimpanzee and human mismatch distributions. Statistically from these data, the human expansion event is as likely to have been driven by global warming at 130 kya as by the Toba eruption at 73.5 kya (see Chapter 1). In general, *post hoc* conclusions about the relationship between specific climatological events and population expansions should be avoided (Rogers and Jorde, 1995). More accurate estimates of nucleotide divergence rates, coupled with data from independent genetic systems, will allow the date of expansion for the human and chimpanzee populations to be more narrowly determined. Data from many sympatric East African taxa should eventually elucidate any broad climatic influences. Major climatic events should have had a significant impact on a wide variety of phylogenetically distinct East African species, if these events were indeed catastrophic in nature (Ruvolo, 1996b).

Chapter 5: Phylogenetic analyses

Introduction

The argument that abiotic change drives evolutionary divergence among taxa is the benchmark of "phylogenetic biogeography," (Brundin, 1988) a discipline tracing its roots to the work of Croizat (1958, 1964). Croizat identified "generalized tracks" or "geosynclines," which he interpreted as representing ancestral distributions of taxa prior to vicariance. This approach was insightful in its integration of geologic process with biotic change, although it lacked quantitative rigor. Croizat's conceptual framework was later combined with the methodology of Hennig (Hennig, 1966) to form "cladistic biogeography," (Humphries *et. al.*, 1988) a discipline striving to reconstruct historical relationships among geographic areas from phylogenies of indigenous taxa, built using character-based methods (Nelson, 1981; Platnick and Nelson, 1984; Rosen and Smith, 1988; Morrone and Carpenter, 1994). The recent availability of molecular data (especially from the mitochondrial genome) has added power to the method by increasing the resolution of reconstructed phylogenies (Avise *et. al.*, 1987; Avise, 1989; Avise, 1994).

The hypothesis that forest taxa in eastern Africa have retracted into and expanded out of Pleistocene refugia (Hamilton, 1981; Kingdon, 1981; Grubb, 1982; Colyn, 1991) implies that present-day forests in this region are related by a unique, if complicated, history. Under the refuge model, forest biota would have differentiated due to: 1) the dispersal of forest taxa into previously unforested areas during post-glacial forest expansion, and 2) the fragmentation and reduction (vicariance) of forests and their indigenous biota during the onset of glacial conditions. Although the relative importance of dispersal and vicariance has been debated (Croizat, 1982; McCoy and Heck, 1983; Brundin, 1988), both forces have probably influenced the present-day distributions of African forest taxa (Hamilton, 1976; Hamilton, 1981; Kingdon, 1981).

The current geographic distribution of chimpanzees corresponds closely to the distribution of forested habitat in Africa (Kortlandt, 1972; Kortlandt, 1983; Goodall, 1986; Teleki, 1989). Although well-adapted to open environments (McGrew, Baldwin and Tutin, 1981; Kortlandt, 1983; Moore, 1992a), chimpanzees are dependent on forest foods (Wrangham *et. al.*, 1991; Wrangham *et. al.*, 1993; Wrangham, Chapman and Chapman, 1994) and on the presence of standing water (Kortlandt, 1983) for survival. Chimpanzees also build and sleep in arboreal nests (Baldwin *et. al.*, 1981; Fruth and Hohmann, 1994a; McGrew, 1994). Nest-building necessitates closed habitat and implies some degree of behavioral adaptation to a forest environment. To the extent that chimpanzees have historically been restricted to forested habitats, evolutionary relationships among extant chimpanzee populations should parallel historical relationships among the forests which they inhabit (see Chapter 1).

The geographic arrangement of locations sampled in the present study, and their proximity to reconstructed Pleistocene refugia, predict a specific pattern of historical relationship for these forested areas. This pattern is presented in Figure 5.1 as a phylogenetic tree (constructed manually). Refuge populations (RIKA and the Zaïrian populations) appear as outgroups relative to the other locations under the assumption that refugia are "source" areas from which the indigenous biota of other peripheral (non-refuge) forests have dispersed (Hamilton, 1976; Hamilton, 1981; Kingdon, 1981; Howard, 1991). Peripheral forests diverge in the order of their geographic proximity to refugia, under the assumption that increasingly-peripheral locations were colonized by forest taxa sequentially during post-glacial forest expansion. This same topology would also be predicted under an alternate vicariant model in which the most peripheral forests were the first to become isolated from a shrinking "mainland" of forest during the onset of glacial conditions (see Chapter 1). Figure 5.1 also reflects the hypothesis that, within eastern forests, southern forests (KUSL, NESN, RIKA) should cluster distinctly from northern forests due to the

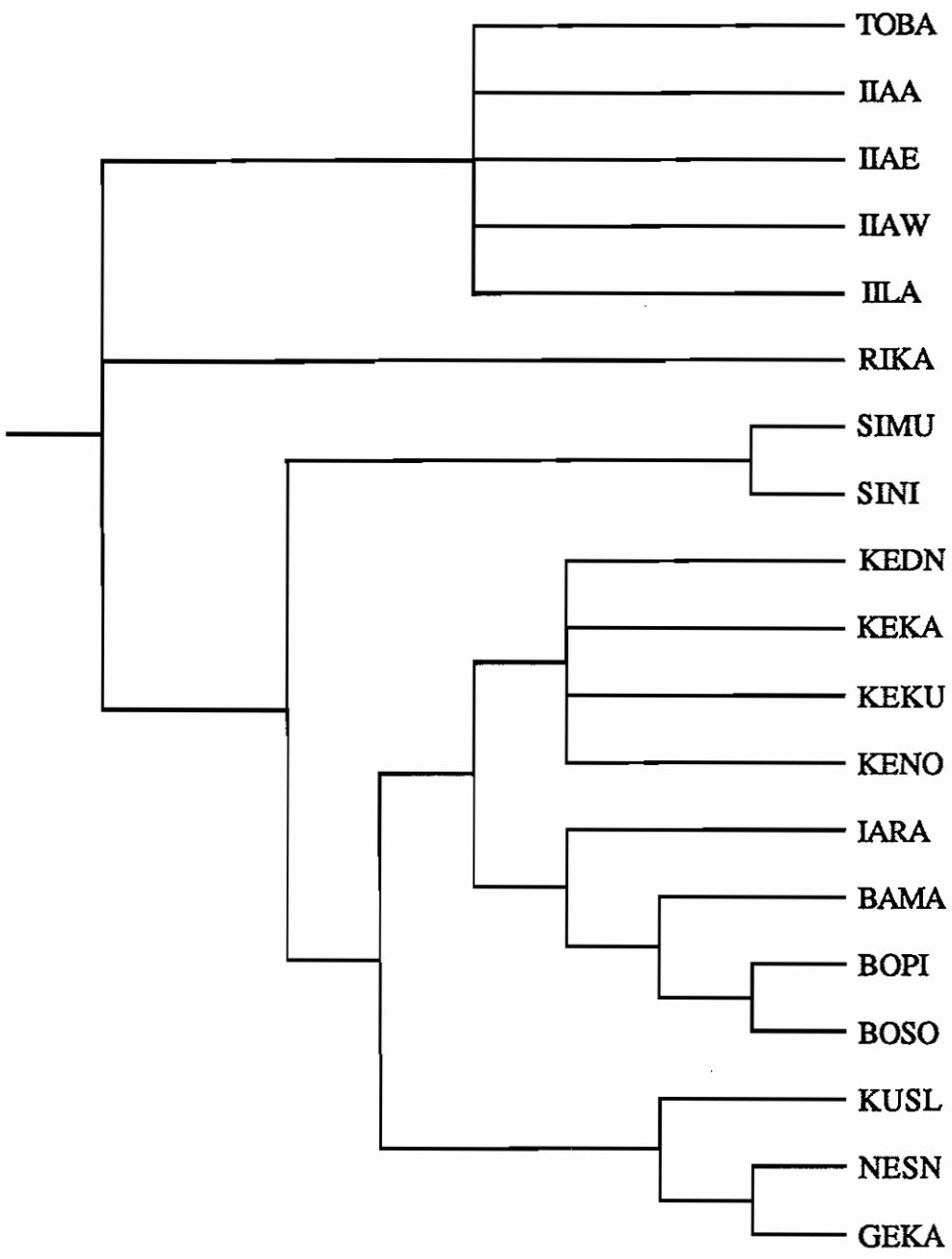


Figure 5.1: Hypothetical phylogeny of sampling locations based on reconstructed locations of Pleistocene forest refugia and the current geographic arrangement of forests (see text for full explanation).

possible role of lakes Edward and George and the Kasinga Channel as a geographic barrier to dispersal.

This chapter investigates the extent to which the refuge model can predict eastern chimpanzee phylogenetic history. Methods of phylogenetic reconstruction are applied to the eastern chimpanzee genetic data at both the haplotypic and population levels. The resulting phylogenies are used to test the hypotheses 1) that chimpanzee populations inhabiting eastern African forests are related by a unique history, and 2) that this history is predictable from the biogeographic relationships of the forests which eastern chimpanzees currently occupy. The first section of the chapter examines evolutionary trees of individual haplotypes; the following sections analyze "population trees" (Templeton, 1993) in which geographic structure is imposed on the data *a priori*.

Haplotype trees

Character-based methods of phylogeny reconstruction are powerful tools for the analysis of evolutionary relationships among DNA sequences. The advantages of the cladistic approach (Hennig, 1966) are evident in its undeniable ability to reconstruct history accurately, which has been demonstrated both theoretically (Penny, 1982; Felsenstein, 1988) and experimentally (Atchley and Fitch, 1991; Hillis *et. al.*, 1992). Nevertheless, non-cladistic (distance-based) methods can also yield reliable phylogenetic information (Nei, Tajima and Tateno, 1983; Saitou and Nei, 1987; Hillis *et. al.*, 1992). Distance (phenetic) methods have the added computational advantage that they arrive at an single reconstruction by an exact algorithm (Swofford and Olsen, 1990). This property obviates the need to search through many potential trees, a process which necessitates heuristic approaches when large numbers of taxa are involved and can be prohibitively time-consuming with very large numbers of taxa (Jin and Nei, 1990; Swofford, 1993; Rice and Donoghue, 1996).

One hundred and twenty three distinct chimpanzee haplotypes were identified in the present study. Of the 368 bp sequenced, 90 positions were phylogenetically informative. The number of unique, bifurcating trees that could be constructed for a data set of this size is astronomical (Felsenstein, 1978), and precludes the use of exact search algorithms which employ maximum-parsimony optimality criteria (Farris, 1986; Felsenstein, 1990; Swofford, 1993). Initial computational attempts using simultaneous runs of PAUP (Swofford, 1993) on separate microcomputers suggest that a statistically meaningful searching of "tree space," even using heuristic approximation, would take on the order of months to years. The problem is analogous to that described for the human mitochondrial gene tree (Vigilant *et. al.*, 1991), from which topological inferences about modern human origins have been problematic due to large numbers of equally-parsimonious cladograms (Maddison, Ruvolo and Swofford, 1992). Furthermore, the number of taxa in the present sample (123) is greater than the number of phylogenetically-informative characters (90), rendering parsimony-based cladograms intrinsically weak (Swofford, 1993; Rogers and Jorde, 1995). For these reasons, the haplotype trees presented below are distance trees. The analysis which follows acknowledges that alternate, more parsimonious trees probably do exist, but assumes that distance methods provide sufficient phylogenetic resolution to address the hypotheses outlined above and in Chapter 1. This assumption is tested below.

Figure 5.2 presents a phylogeny of eastern chimpanzee haplotypes reconstructed from a matrix of distances generated using the DNAdist module of the computer program PHYLIP with a maximum-likelihood distance correction and a transition:transversion ratio of 10:1. Transition-transversion ratios as low as 1:1 and as high as 30:1 were also used, but did not alter the results of this or the following analyses. The algorithm used to generate the tree is the UPGMA (unweighted pair group method using arithmetic averages) algorithm (Sokal and Michener, 1958). The tree is rooted using outgroup sequences from three central African chimpanzees (*P. t. troglodytes*), three west African chimpanzees (*P. t. verus*), two bonobos (*P. paniscus*) and two humans (*Homo sapiens*). One west

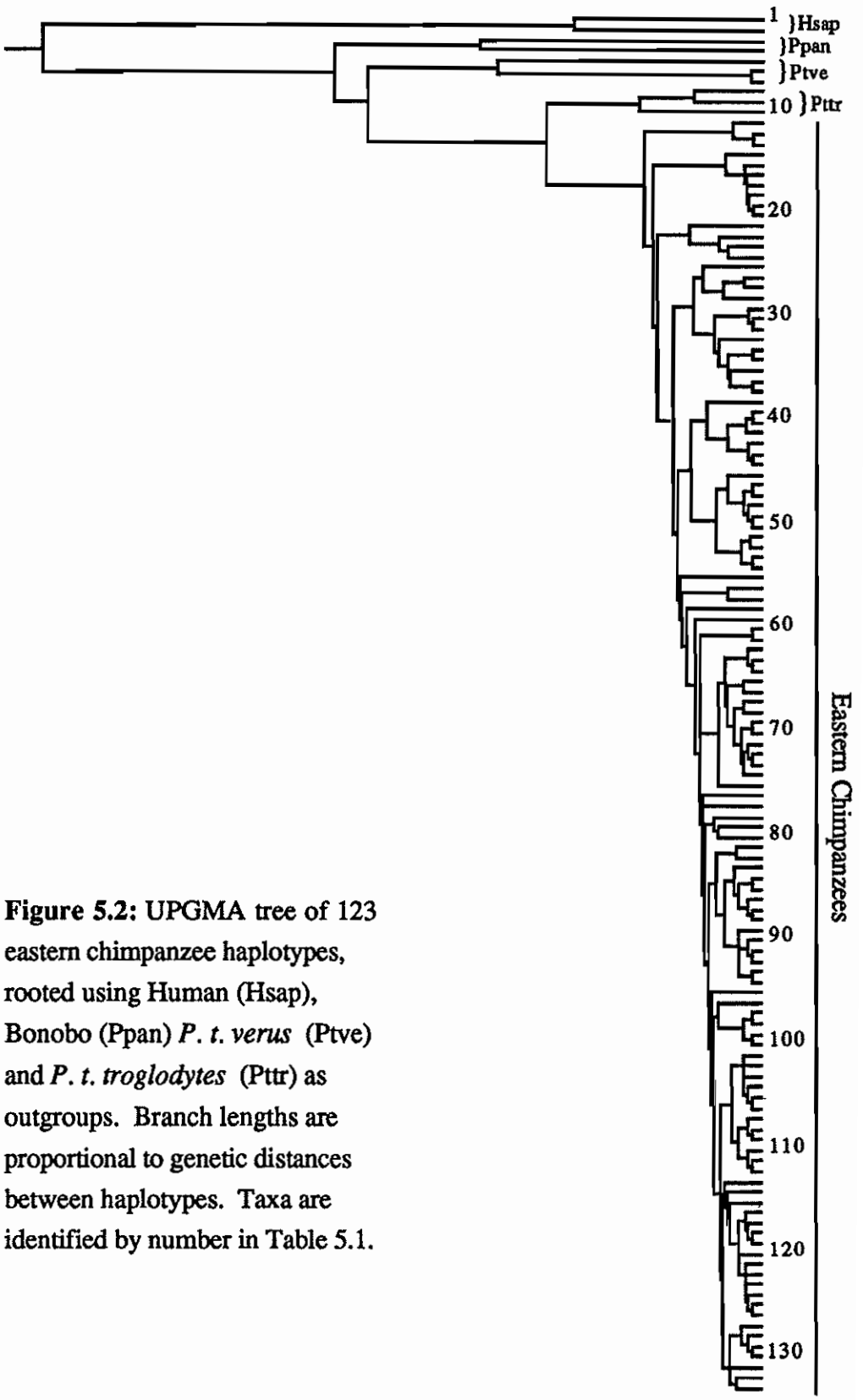


Figure 5.2: UPGMA tree of 123 eastern chimpanzee haplotypes, rooted using Human (Hsap), Bonobo (Ppan) *P. t. verus* (Ptve) and *P. t. troglodytes* (Pttr) as outgroups. Branch lengths are proportional to genetic distances between haplotypes. Taxa are identified by number in Table 5.1.

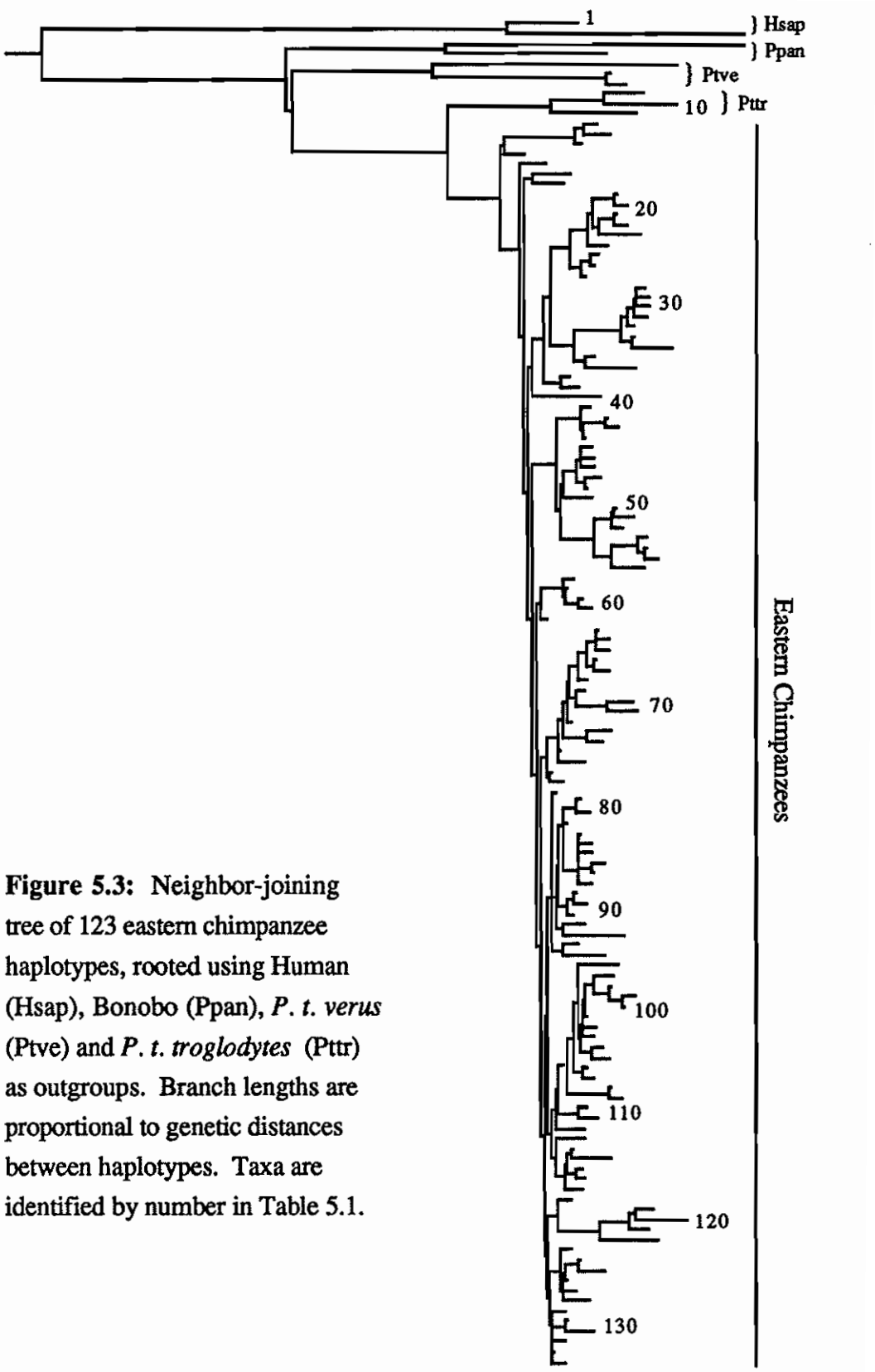


Figure 5.3: Neighbor-joining tree of 123 eastern chimpanzee haplotypes, rooted using Human (Hsap), Bonobo (Ppan), *P. t. verus* (Ptve) and *P. t. troglodytes* (Pttr) as outgroups. Branch lengths are proportional to genetic distances between haplotypes. Taxa are identified by number in Table 5.1.

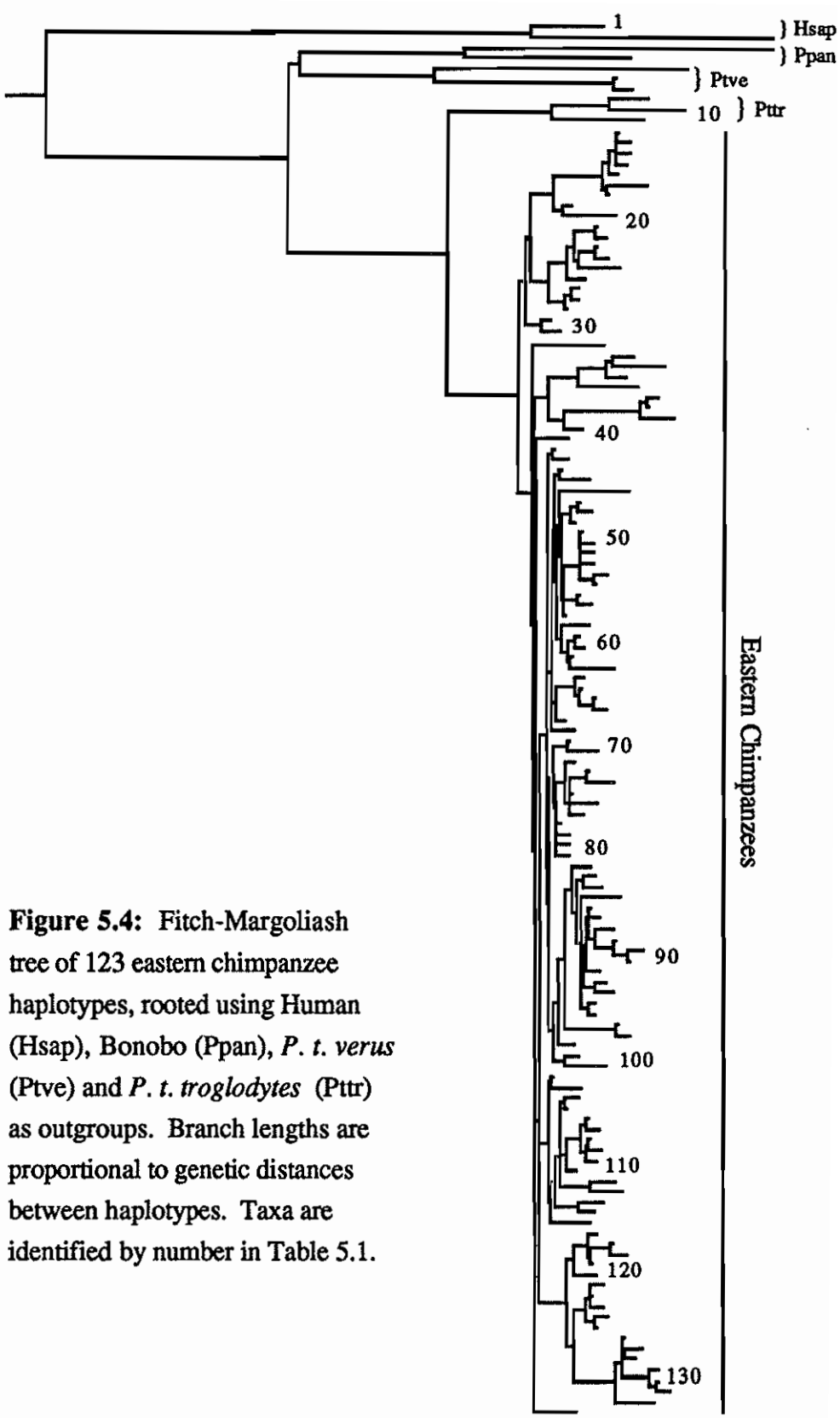


Figure 5.4: Fitch-Margoliash tree of 123 eastern chimpanzee haplotypes, rooted using Human (Hsap), Bonobo (Ppan), *P. t. verus* (Ptve) and *P. t. troglodytes* (Pttr) as outgroups. Branch lengths are proportional to genetic distances between haplotypes. Taxa are identified by number in Table 5.1.

Table 5.1: Identification of taxa shown on phylogenetic trees in Figures 5.2, 5.3, and 5.4

Taxon number*	Haplotype identification**		
	UPGMA	Neighbor-Joining	Fitch-Margoliash
1	Homo01	Homo01	Homo01
2	Homo02	Homo02	Homo02
3	Ppan01	Ppan01	Ppan01
4	Ppan02	Ppan02	Ppan02
5	Ptve01	Ptve01	Ptve01
6	Ptve02	Ptve02	Ptve02
7	Ptve03	Ptve03	Ptve03
8	Pttr01	Pttr01	Pttr01
9	Pttr02	Pttr02	Pttr02
10	Pttr03	Pttr03	Pttr03
11	1	3	45
12	2	1	14
13	3	2	12
14	4	68	13
15	5	86	11
16	6	70	3
17	7	69	2
18	8	27	1
19	9	26	68
20	10	24	86
21	11	23	78
22	12	22	76
23	13	25	116
24	14	21	103
25	15	20	48
26	16	19	99
27	17	10	98
28	18	7	97
29	19	6	96
30	20	8	93
31	21	9	91
32	22	5	92
33	23	4	94
34	24	18	95
35	25	15	102
36	26	17	100
37	27	16	101
38	28	45	121
39	29	44	118
40	30	42	119
41	31	41	120
42	32	43	66
43	33	40	90
44	34	38	89
45	35	39	88
46	36	37	87
47	37	36	110
48	38	35	117

Table 5.1 (continued): Identification of taxa shown on phylogenetic trees in Figures 5.2, 5.3, and 5.4

	Haplotype identification**		
*Taxon number**	UPGMA	Neighbor-Joining	Fitch-Margoliash
49	39	34	112
50	40	32	85
51	41	33	109
52	42	31	107
53	43	30	49
54	44	29	108
55	45	28	105
56	46	90	106
57	47	89	115
58	48	88	114
59	49	87	111
60	50	110	113
61	51	84	122
62	52	83	56
63	53	81	55
64	54	82	65
65	55	80	62
66	56	79	61
67	57	74	64
68	58	47	54
69	59	46	52
70	60	75	53
71	61	72	63
72	62	71	58
73	63	77	57
74	64	73	60
75	65	78	59
76	66	76	51
77	67	116	50
78	68	102	123
79	69	100	104
80	70	101	67
81	71	96	77
82	72	93	73
83	73	91	74
84	74	94	75
85	75	95	82
86	76	92	80
87	77	99	84
88	78	98	83
89	79	97	81
90	80	103	79
91	81	48	47
92	82	104	46
93	83	67	72
94	84	65	71
95	85	64	70
96	86	54	44
97	87	52	42

Table 5.1 (continued): Identification of taxa shown on phylogenetic trees in Figures 5.2, 5.3, and 5.4

	Haplotype identification**		
*Taxon number**	UPGMA	Neighbor-Joining	Fitch-Margoliash
98	88	53	41
99	89	62	43
100	90	63	35
101	91	61	39
102	92	58	40
103	93	57	38
104	94	60	37
105	95	59	15
106	96	51	34
107	97	50	32
108	98	56	33
109	99	55	31
110	100	122	30
111	101	121	29
112	102	120	28
113	103	66	69
114	104	118	10
115	105	119	7
116	106	117	6
117	107	123	8
118	108	14	9
119	109	12	4
120	110	13	5
121	111	11	18
122	112	109	36
123	113	107	27
124	114	49	26
125	115	108	24
126	116	106	23
127	117	105	22
128	118	114	25
129	119	112	21
130	120	85	20
131	121	113	19
132	122	115	17
133	123	111	16

* Taxon numbers are sequential, and are ordered from the top to the bottom of each tree (see Figures 5.2-5.4).

* Haplotype identification numbers refer to Appendices 2 and 3

African chimpanzee sequence and both bonobo sequences were taken from the literature (Horai *et. al.*, 1992; Morin *et. al.*, 1994a), as were the two human sequences (Vigilant *et. al.*, 1991), which were specifically chosen to be maximally-divergent within Vigilant's 135 haplotype world human sample. Taxa are labeled with numbers, sequentially from the top of the tree to the bottom. Each of these numbers refers to a specific haplotype. The haplotypes corresponding to these numbers are identified in Table 5.1. The haplotypes themselves (raw DNA sequences) are presented in Appendix 2, and the locations and population frequencies of each haplotype are given in Appendix 3. Haplotypes on all following trees are indexed in the same way in Table 5.1.

Diversity within eastern chimpanzees (measured as the length of the branches connecting the most divergent eastern chimpanzee sequences) is lower than that within humans. Specifically, the two human sequences differ from one another by a total of 22 substitutions (20 transitions and 2 transversions). Correcting these distances for multiple substitutions with a maximum-likelihood distance correction and a transition: transversion ratio of 10:1 yields a distance of 23 substitutions (21 transitions and 2 transversions) between the human haplotypes. In contrast, the most divergent clade of eastern chimpanzee haplotypes (represented by haplotypes 1-3 on the UPGMA tree) differs from the other eastern chimpanzee haplotypes by a maximum of 20 substitutions, 18 of which are transitions and 2 of which are transversions. Maximum-likelihood distance correction and a transition: transversion ratio of 10:1 changed this estimate to 21 substitutions (19 transitions and 2 transversions). The approximate coalescent time for eastern chimpanzees would therefore be (using corrected estimates) 21/23, or 91% of the human estimate. A human estimate of 298 kya (Ruvolo *et. al.*, 1993) would therefore imply a coalescence time of 272 kya for the last common ancestor of all eastern chimpanzees.

The near identity of the corrected and uncorrected distance estimates given above suggests that the first hypervariable region of d-loop evolves at a rate appropriate for

discerning evolutionary relationships on the order of tens of thousands of years. In contrast, the uncorrected genetic distance between eastern chimpanzees and humans was 65 substitutions (61 transitions, 4 transversions), which changed to 233 substitutions (221 transitions, 12 transversions) when a maximum-likelihood correction and a transition: transversion ratio of 10:1 were used. Since the length of the entire region sequenced was only 368 bp, this observation suggests that almost every site has undergone substitution and that d-loop has become "saturated" with substitutions since the divergence of humans and chimpanzees (approximately six million years ago). Therefore, this region of mitochondrial DNA will not likely yield accurate dates for events on this time scale (Kocher and Wilson, 1991; Ruvolo *et. al.*, 1993).

Diversity within eastern chimpanzees is also lower than that within western African chimpanzees (*P. t. verus*; uncorrected maximum difference: 28 transitions, 2 transversions; corrected maximum difference: 29 transitions, 3 transversions) and bonobos (*P. paniscus*; uncorrected maximum difference: 29 transitions, 2 transversions; corrected maximum difference: 30 transitions, 3 transversions), but is slightly higher than that within the central African chimpanzees sampled (*P. t. troglodytes*; uncorrected maximum difference: 14 transitions, 1 transversion; corrected maximum difference: 15 transitions, 1 transversion). Diversity within bonobos and west African chimpanzees therefore exceeds that within humans, despite minimal geographic sampling within these subspecies. Indeed, maximum sequence difference within *P. t. verus* and *P. paniscus* is greater even than the genetic distance between *P. t. schweinfurthii* and *P. t. troglodytes* (uncorrected distance: 24 transitions, 2 transversions; corrected distance: 25 transitions, 2 transversions).

Within eastern chimpanzees, haplotypes tend not to form exclusive regional clades on the phylogenetic tree. Most major clades consist of sequences from both Zaïrian forests and eastern forests. However, the most divergent clade (from the others) within eastern chimpanzees consists exclusively of sequences from Zaïre (Figure 5.2, sequences 1-3, all

found in the IIAA population). This topology is thus consistent with a Zaïrian location for the basal node linking eastern chimpanzees. Sampling locations also do not form exclusive clades on this tree. Lack of location-level clustering would be expected from previous analyses which suggest that considerable gene flow has occurred across the landscape and throughout the history of the subspecies (see Chapter 3). Entebbe Zoo haplotypes, the geographic provenances of which are largely unknown, all cluster within the *P. t. schweinfurthii* clade, albeit in different subclades.

Figure 5.3 presents a haplotype tree (inferred from the same distance matrix as the UPGMA tree) constructed using the neighbor-joining algorithm of Saitou and Nei (1987), The tree was outgroup-rooted with the input order of ingroup taxa randomized. This algorithm has the advantage that it does not assume equal rates of evolution among lineages. It has the disadvantage that, because the resulting tree is additive, negative branch lengths are possible and allowed (Swofford and Olsen, 1990). Negative branch lengths did not, however, occur in this case, although they were not specifically excluded. The basic properties of the neighbor-joining tree are identical to those of the UPGMA tree. Relationships among and levels of diversity within outgroup taxa are the same, and reinforce the notion that eastern chimpanzees are at the low end of the genetic diversity spectrum within the hominoid subspecies, including humans. Obvious regional and population-level subdivisioning is, again, absent. A maximally-divergent cluster of exclusively Zaïrian sequences appears as an outgroup to the rest of eastern chimpanzees, consistent with the possibility of a Zaïrian location for the ancestral sequence at the basal node linking *P. t. schweinfurthii.* Three of these four sequences are the same as those observed in the most divergent UPGMA cluster, and the fourth is from the TOBA population (the westernmost Zaïrian location sampled). Most major clades within eastern chimpanzees are preserved between the UPGMA and neighbor-joining trees (see Table 5.1), although relationships among clades vary. Again, Entebbe Zoo haplotypes all fall within the eastern chimpanzee clade.

Figure 5.4 presents a haplotype tree inferred using the algorithm of Fitch and Margoliash (Fitch and Margoliash, 1967). Like neighbor-joining, the Fitch-Margoliash algorithm creates an additive tree in which non-equal branch lengths are allowed. However, the Fitch-Margoliash method does not allow negative branch lengths and thus may avoid difficulties associated with their interpretation. The Fitch-Margoliash tree is roughly similar to the UPGMA and neighbor-joining trees in that it shows no evidence of phylogenetic partitioning on a regional or population level. Also, differences in relative levels of diversity within subspecies are maintained. However, the tree is topologically different in its clustering of *P. t. verus* with *P. paniscus* to the exclusion of *P. t. troglodytes* and *P. t. schweinfurthii*. This evolutionary scenario is unlikely, and probably results from short internodal branch lengths at this phylogenetic depth, in addition to homoplasy due to the rapid rate of evolution of the gene region. The topology of the Fitch-Margoliash tree is also different in its lack of a maximally-divergent clade of exclusively Zaïrian sequences within eastern chimpanzees. Although the same haplotypes found in the maximally-divergent clades of the UPGMA and neighbor-joining trees do cluster in the maximally-divergent Fitch-Margoliash clade, they are interspersed with haplotypes from a variety of eastern and Zaïrian locations (Table 5.1). Other clades observed in the UPGMA and neighbor-joining trees do not generally occur in the Fitch-Margoliash tree (see Table 5.1). However, the Fitch-Margoliash tree preserves the clustering of all Entebbe Zoo chimpanzees with *P. t. schweinfurthii*.

Statistical analyses of haplotype trees

Slatkin and Maddison (1989) describe a tree-based method for detecting restrictions to gene flow among populations within a species. Given a phylogeny of haplotypes (the geographic origins of which are known) one can reconstruct the minimum number of interpopulational migration events needed to explain the spatial distribution of haplotypes.

In practice, location is coded as an unordered, multistate character and its evolution is traced along a tree reconstructed from non-geographic (genetic) information (Slatkin and Maddison, 1989; Maddison and Maddison, 1992). The minimum number of character state changes reconstructed along the tree represents the minimum number of inferred migration events, *s*. Non-parametric 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 taxon number. If observed numbers of migration events 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 was applied to each of the haplotype trees described above. Three multistate, unordered geographic characters were defined. The first was a 19-state character representing each of the geographically-defined sampling locations in the study (EEZO excluded). The second was a two-state unordered character representing region (eastern forests versus Zaïrian forests). The third "geographic" character was a two-state unordered character used to differentiate sequences found in the captive Entebbe Zoo population from sequences not present in this population. Results are shown in Figure 5.5. Histograms (open bars) represent the distribution of *s* values generated by tracing each character along 1000 randomly-generated trees using the computer program MacClade (Maddison and Maddison, 1992). Empirical values of *s* inferred from UPGMA, neighbor-joining and Fitch-Margoliash trees (shown in Figures 5.1, 5.2 and 5.3) are indicated by letters (U, N, F, respectively) on each panel.

For the location character, observed *s* values are lower than all values generated for the simulated null distribution, indicating that gene flow within the subspecies is significantly restricted at the population level (Figure 5.5, upper panel; $p < 0.001$). Gene flow has similarly been restricted at the regional level, between Zaïrian and eastern forests (Figure 5.5, middle panel; $p < 0.001$). These results reconfirm those of non-tree-based

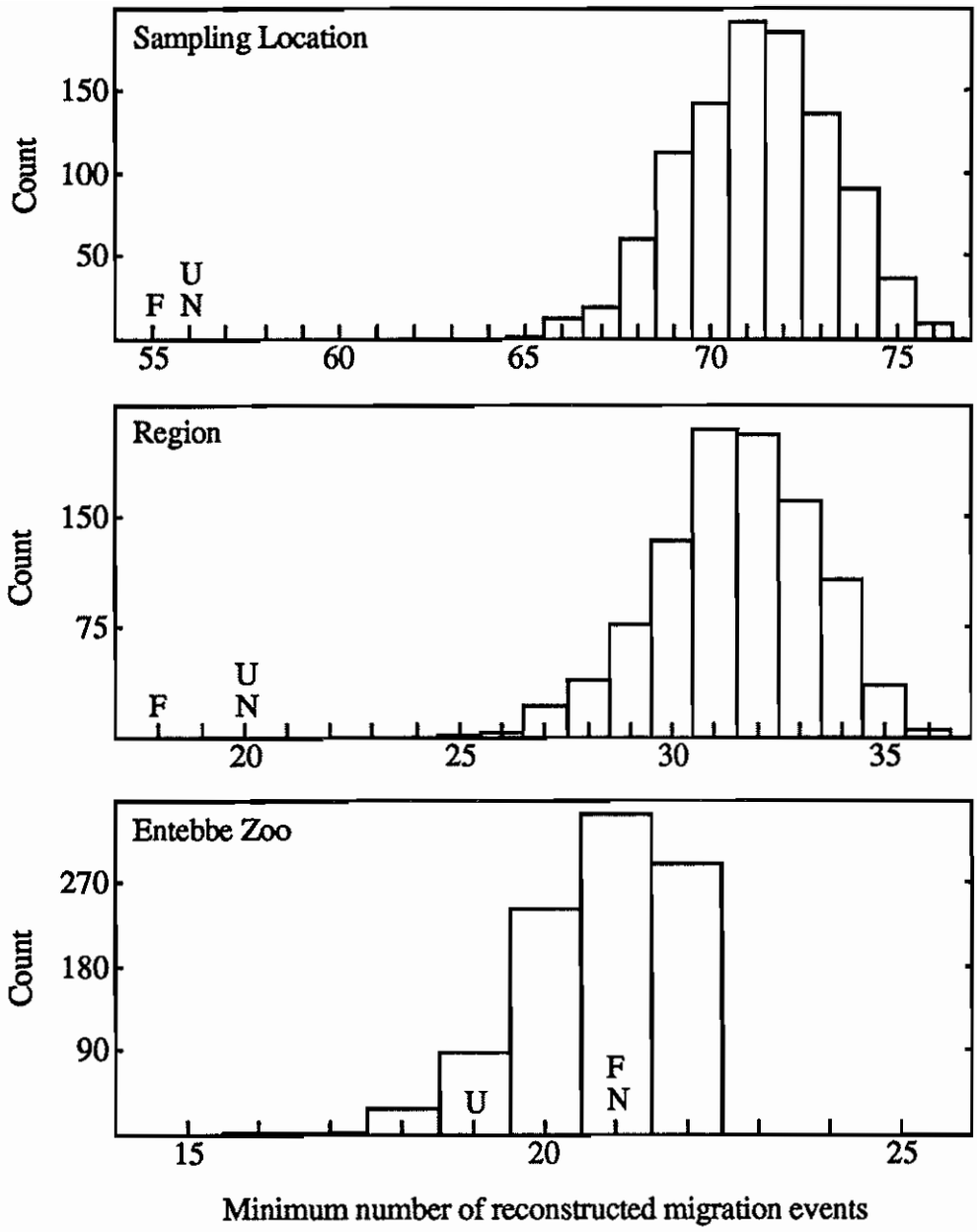


Figure 5.5: Null distributions of cladistically-reconstructed migration events based on randomly-generated trees of 123 haplotypes. For each haplotype, sampling location, region (Eastern vs. Zaïrian forests), and Entebbe Zoo were coded as multistate unordered characters, which were traced on 1000 random trees constructed using MacClade (Maddison and Maddison, 1992). Eurytopic haplotypes were excluded from the analysis. Numbers of reconstructed changes in these characters represent minimum numbers of inferred migration events (Slatkin and Maddison, 1989). Letters indicate numbers of migration events inferred for trees of haplotypes constructed using the Fitch-Margoliash (F), UPGMA (U) and neighbor-joining (N) methods.

techniques described previously (see Chapter 3), which also detected restricted (but not severely so) gene flow at the population level. This analysis demonstrates the importance of using quantitative statistical techniques to determine the extent of population subdivisioning implied by phylogenetic trees. Qualitatively, none of the haplotype trees presented in Figures 5.2 - 5.4 show obvious geographic partitioning. Even though it is pronounced, restricted gene flow could not have been detected by visual inspection.

In the case of the Entebbe Zoo character, "migration events" represent human-induced gene-flow into and out of Entebbe Zoo, resulting from the illegal capture and transport, and ultimate confiscation, of animals from local poachers, international traders, and circuses. Analysis of the Entebbe Zoo character (Figure 5.5, lower panel) demonstrates 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. Entebbe Zoo population is therefore unlike any natural population. Entebbe Zoo chimpanzees do not tend to originate from any single geographic location. Rather, poaching is a problem of cosmopolitan proportion.

The haplotype phylogenies shown in Figures 5.2 - 5.4 were also used to investigate molecular evolutionary processes within hypervariable region 1 of the mitochondrial control region (Kocher and Wilson, 1991). Character-state changes (in this case, nucleotide base substitutions) were traced along UPGMA, neighbor-joining and Fitch-Margoliash trees of haplotypes for both the eastern chimpanzee sample (n = 123) and Vigilant's (1991) world human sample (n = 135) using the computer program MacClade (Maddison and Maddison, 1992). Results are presented in Figure 5.6 for the neighbor-joining tree only, since results for the UPGMA and Fitch-Margoliash trees were indistinguishable. Results obtained by non-tree-based methods (*i.e.* simple counts of nucleotide variation along the sequence) also yielded a pattern statistically indistinguishable

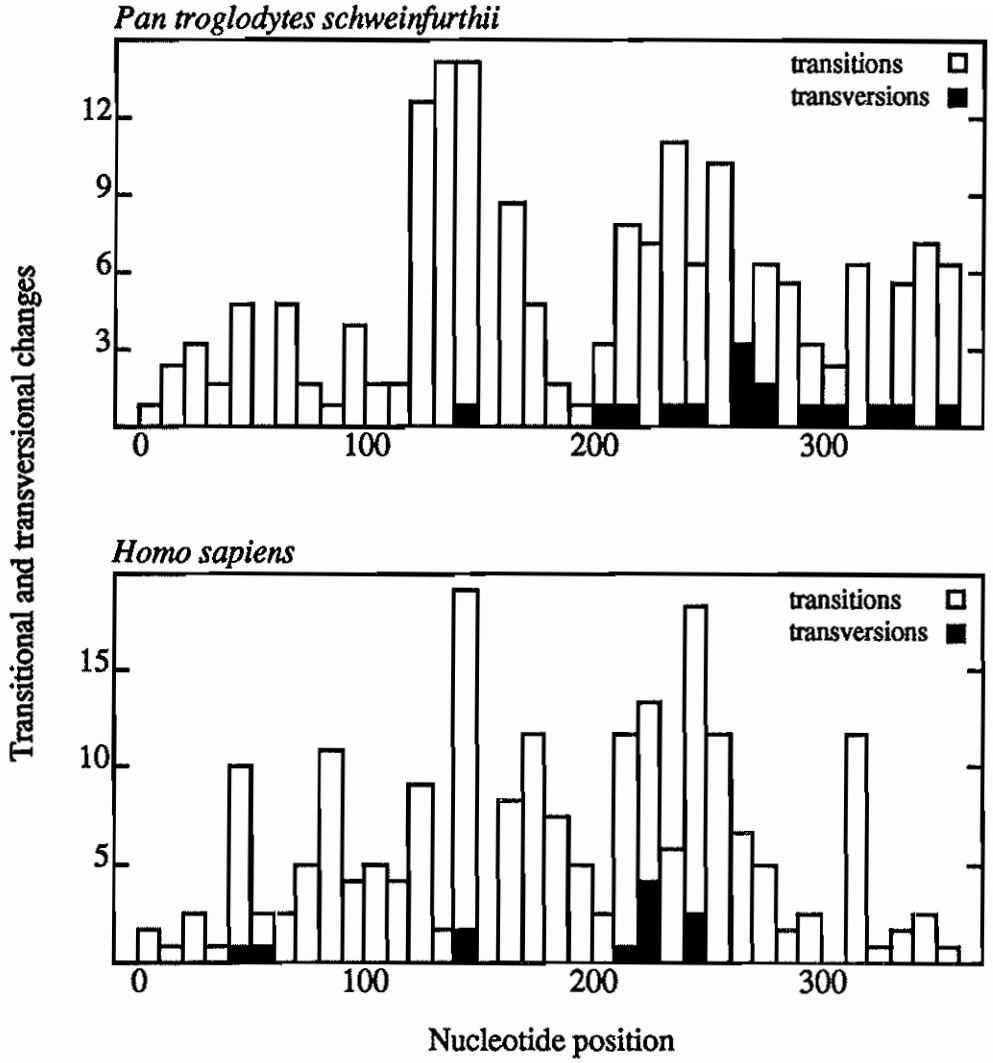


Figure 5.6: Comparative molecular evolution of eastern chimpanzee (n = 123) and human (n = 135) mitochondrial d-loop first hypervariable regions. Minimum numbers of transitional and transversional changes were calculated at 10-nucleotide intervals along the gene region sequenced (hypervariable region 1) using the program MacClade (Maddison and Maddison, 1992). Changes were traced along a phylogenetic tree of haplotypes constructed for each subspecies with the program PHYLIP (Felsenstein, 1993) using the neighbor-joining algorithm (Saitou and Nei, 1987). Results were indistinguishable when trees constructed using other methods were analyzed. Human data were taken from Vigilant *et. al.* (1991).

from the one shown in Figure 5.6. Inferred numbers of transitional changes (open bars) and transversional changes (closed bars) are plotted at ten-nucleotide intervals along the gene region, which corresponds to bases 16041 - 16413 of human reference sequence (Anderson *et. al.*, 1981).

Patterns of nucleotide substitution are similar between humans and eastern chimpanzees. Numbers of transitional changes are greatest in the middle of the region sequenced, as predicted from interspecific comparisons (Kocher and Wilson, 1991). The transitional bias observed in the eastern chimpanzee data (approximately 19:1) is similar to that observed for the human data (approximately 17:1). In neither taxon were the numbers of transitional or transversional changes normally distributed (Kolmogorov-Smirnov test using standard normal distribution; $p < 0.0001$ in all cases), nor could the distribution of changes be made to fit a normal distribution by standardizing the mean and standard deviation to zero and one, respectively (Lillefors test, $p < 0.0001$ in all cases; Lillefors, 1967). Therefore, the statistical comparisons presented below are nonparametric. Tests were run using the program SYSTAT (Wilkinson, Hill and Vang, 1992).

In eastern chimpanzees, areas of high transitional variation did not tend to cluster along the length of the sequence (Wald-Wolfowitz runs test; $p = 0.41$). Similarly, transversional changes were not clustered along the gene region in eastern chimps (Wald-Wolfowitz runs test; $p = 0.71$). In humans, however, transitional changes were clustered (Wald-Wolfowitz runs test; $p = 0.01$), although transversional changes were not (Wald-Wolfowitz runs test; $p = 0.15$). Direct comparisons of human and eastern chimpanzee sequences detected no statistical difference, however, either in the case of transitions (Wilcoxon signed rank test; $p = 0.40$), or in the case of transversions (Wilcoxon signed rank test; $p = 0.97$). No evidence therefore exists that the molecular evolutionary dynamics of this segment of the mitochondrial control region have changed since the divergence of humans and chimpanzees.

The analyses presented above operate under the previously-stated assumption that

distance methods of tree reconstruction yield phylogenetically-meaningful topologies. To the extent that parsimony is the most appropriate optimality criterion for reconstructing phylogeny (Farris, 1983; Sober, 1988), this assumption may be tested empirically. Using MacClade (Maddison and Maddison, 1992), 1000 randomly-generated equiprobable trees were generated for the full data set of 123 chimpanzee haplotypes and 10 outgroups. The length of each tree was determined from the molecular data to generate a null distribution of tree lengths. Lengths were then calculated for each of the distance trees described above. The UPGMA (length = 442), neighbor-joining (length = 408) and Fitch-Margoliash (length = 414) trees were all significantly shorter than the shortest random tree (length = 905), indicating that these distance methods do, in fact, arrive at significantly "better" topologies than would be expected by chance.

Population trees

When analyzed using MacClade (Maddison and Maddison, 1992), trees of haplotypes in which populations were constrained to be monophyletic were considerably less parsimonious than any of the reconstructed distance trees presented above (length of shortest population monophyly constraint tree = 522). This observation does not, however, imply that the populations lack a unique, recoverable history. Gene flow among eastern chimpanzee populations has been extensive, but may not have entirely obscured all genetic record of the order in which the populations were founded, either through dispersal after glacial retreats, or through the vicariance of chimpanzee habitat subsequent to the onset of glaciation. The following analyses impose a population structure on the data *a priori* in an effort to reconstruct such a history.

Distance matrices were generated for the 19 geographically-defined populations in the study using frequencies of the 123 identified alleles. All distances were computed using the program PHYLIP (Felsenstein, 1990). Distances were converted to trees using

the Fitch-Margoliash algorithm available in PHYLIP. Other tree-building methods (UPGMA, neighbor-joining) yielded virtually identical phylogenies. Results are presented in Figure 5.7. The trees are unrooted, since midpoint rooting tended to cause visual biases (overestimation) of the degree to which populations formed clusters. For purposes of presentation, branch lengths were not drawn proportional to distance, although relative branch lengths are indicated above branches.

Three allelic distance measures available in PHYLIP were used. The first (top panel), that of Cavalli-Sforza and Edwards (1967), is:

$$D^2 = 4 \sum_m [\, 1 - \sum_i p_{1mi}^{1/2} p_{2mi}^{1/2} \,] / \sum_m (a_m - 1)$$

where i alleles occur at m loci, and a is the number of alleles at the mth locus, and where p_{1mi} is the frequency of the ith allele at locus m. This arc-distance measure is similar to that of Reynolds, Weir and Cockerham (1983; bottom panel) in its assumption that differences among populations result from drift alone. The Reynolds, Weir and Cockerham distance is:

$$D^2 = \frac{\sum_m \sum_i (p_{1mi} - p_{2mi})^2}{2 \sum_m (1 - \sum_i p_{1mi} p_{2mi})}$$

where m is summed over loci and i over alleles at the mth locus, and where p_{1mi} is the frequency of the ith allele at locus m. The Cavalli-Sforza and Edwards measure is different (and perhaps superior) in its robusticity to biases introduced by within-population heterozygosity (Swofford and Olsen, 1990). Nei's (1972) genetic distance (middle panel) is:

$$D = -\ln \left[\frac{\sum_m \sum_i p_{1mi} p_{2mi}}{(\sum_m \sum_i p_{1mi}^2)^{1/2} (\sum_m \sum_i p_{2mi}^2)^{1/2}} \right]$$

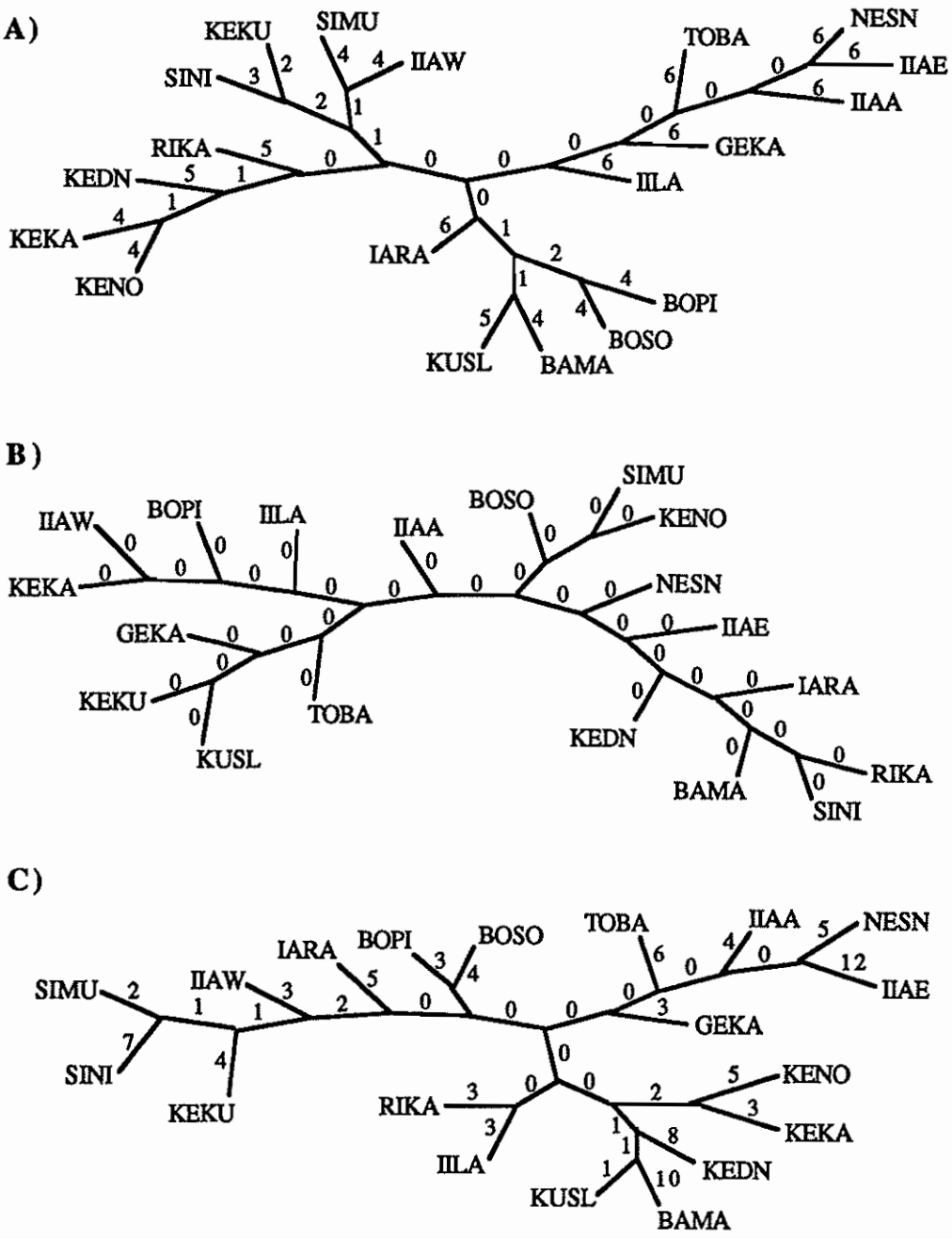


Figure 5.7: Relationships among populations based on allele frequencies. Three genetic distance matrices were calculated using the program PHYLIP (Felsenstein, 1993). Distance measures were those of A) Cavalli-Sforza and Edwards (1967), B) Nei (1972), and C) Reynolds, Weir and Cockerham (1983). Trees were constructed using PHYLIP with the algorithm of Fitch and Margoliash (1967). Topologies were identical when other tree-building methods were used. Numbers above branches (not drawn to scale) are relative branch lengths.

where the notation is the same as above. It differs from the other two measures in that it considers mutation an important factor in population differentiation. This assumption is likely to be valid in the present case, since the clustering of similar alleles in the haplotype trees described above implies the localized evolution of unique alleles through point mutation.

These distance measures are unlikely to have reconstructed population history accurately in the present case. Nei, Tajima and Tateno (1983) demonstrated by computer simulation that accurate phylogenetic reconstructions from gene frequency data were possible only when 30 or more loci were used, even when a maximally-efficient tree reconstruction algorithm was employed and only eight taxa were considered. For the present number of loci (one) and taxa (19), the likelihood that a correct topology was reconstructed is effectively zero. The distance approach with gene frequency data should thus be viewed cautiously. It is potentially useful only for identifying particularly robust population groupings.

In the Cavalli-Sforza and Edwards (1967) and Reynolds, Weir and Cockerham (1983) trees, locations within Budongo Forest (BOPI and BOSO) tend to cluster, as do locations within Semliki Forest (SIMU and SINI). In all three trees, Kibale Forest locations (KEDN, KEKA, KEKU, KENO) are paraphyletic, although KEDN, KEKA and KENO do tend to cluster on the Cavalli-Sforza and Edwards (1967) and Reynolds, Weir and Cockerham (1983) trees. Few other strong similarities exist among the three trees. Branch lengths in the Nei (1972) tree are all approximately equal (zero), suggesting a "star phylogeny" with little resolution. A similar pattern is also suggested by the Cavalli-Sforza and Edwards (1967) and Reynolds, Weir and Cockerham (1983) trees; in both trees, central nodes are linked by short branch lengths while long branches appear only for terminal taxa.

The number of populations in this study (19) is considerably smaller than the number of haplotypes (123), which makes cladistic analyses using parsimony criteria

possible at the population level. The analysis which follows attempts to reconstruct "ancestral" alleles for each sampling location using cladistic inference; parsimony criteria are then used to reconstruct phylogenetic relationships among these alleles (Rice and Donoghue, 1996). The analysis considers only endemic alleles (alleles found exclusively in a single sampling location) under the assumption that they evolved locally from a unique ancestral sequence which was also geographically localized. This analysis further assumes that reconstructed ancestral alleles represent "founder" alleles from the initial population inhabiting each sampling location. Eurytopic alleles (alleles shared among sampling locations) were excluded under the assumption that they represent recent migration and could confound phylogenetic reconstruction by introducing homoplasy.

For each population, a maximum-parsimony tree of endemic haplotypes was generated using the branch and bound algorithm of PAUP (Swofford, 1993) and a transition:transversion ratio of 10:1. This algorithm is an exact search method, practical only for numbers of taxa less than approximately 13. One central African chimpanzee, one west African chimpanzee and one bonobo sequence were included in each search as outgroups. In 17 of 19 cases, a single most parsimonious tree was identified. In the remaining 2 cases, more than one (2 and 4, respectively) equally-parsimonious tree was identified. PAUP was then used to reconstruct the ancestral state of each of the 368 characters (nucleotide positions) at the basal node linking the eastern chimpanzee haplotypes in each tree. Ancestral haplotypes were input into PAUP along with sequences from a human, a bonobo, a western African chimpanzee, and a central African chimpanzee. PAUP was used with the heuristic search option (10 random-addition replicates) and a transition:transversion ratio of 10:1 to search for maximally parsimonious trees (transition:transversion ratios of 1:1 and 30:1 were also tested, but did not alter the results of the analysis). One thousand bootstrap replicates of the data (Felsenstein, 1985) were run to estimate statistical confidences around reconstructed groupings.

Results are presented in Figure 5.8. Because 1163 equally-parsimonious trees were found during the heuristic search, the cladogram presented is a 50% majority-rule bootstrap consensus tree in which all possible groupings were allowed. Although outgroup relationships were reconstructed with high confidence, ingroup relationships were not. Bootstrap values within the eastern chimpanzee clade were universally low and internal branch lengths (minimum numbers of reconstructed changes) were all ≤ 1. The topology is also inconsistent with the predictions of Pleistocene refuge theory, as outlined at the beginning of the chapter. The hypothetical population tree presented in Figure 5.1 was a considerably less-parsimonious explanation of the data (length = 152) than the reconstructed tree (length = 139), when analyzed using MacClade (Maddison and Maddison, 1992). "Refuge" locations (Zaïrian locations and RIKA) are dispersed throughout the tree in Figure 5.8, rather than being outgroups, as predicted if other populations were founded by chimpanzees dispersing from them (see Figure 5.1). BAMA (Bugoma), a peripheral forest which probably did not exist in its present location before approximately 200 years ago (Howard, 1991) is reconstructed as an outgroup to all other populations. Locations within Kibale Forest are paraphyletic, as are locations within Semliki Forest. The two Budongo Forest locations do cluster, but with only 4% bootstrap confidence.

Eurytopic haplotypes were excluded from the analysis described above because they were assumed to be represent recent migration. However, eurytopic haplotypes may be good indicators of relatedness between populations under the alternate assumption that they reflect recent common ancestry. This would be true under a model of population origin in which founder populations dispersing from a common ancestral population shared greater numbers of alleles than founding populations dispersing from different ancestral populations. This would also be true in the case of vicariance, under the assumption that recently-vicariated populations retain greater numbers of alleles in common than do

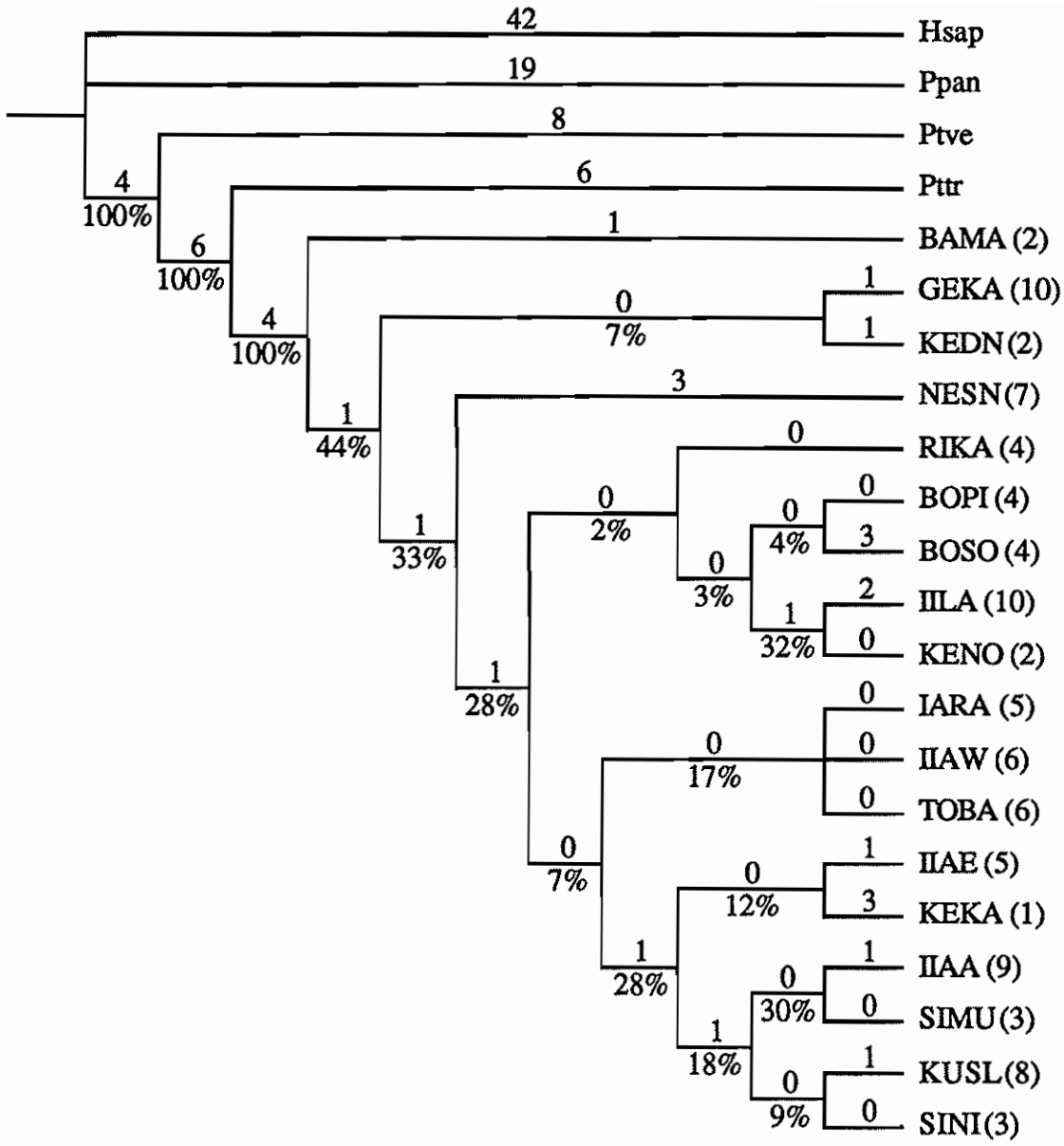


Figure 5.8: Maximum parsimony tree of ancestral haplotypes. The cladogram is a 50% majority-rule consensus tree computed from 1000 bootstrap replicates of the data using the heuristic search option in PAUP (Swofford, 1993). Numbers above branches are minimum numbers of inferred changes. Numbers below branches are bootstrap values. The tree is rooted using Human (Hsap), Bonobo (Ppan), *P. t. verus* (Ptve) and *P. t. troglodytes* (Pttr) as outgroups. Tree length = 139; rescaled consistency index = 0.588. Ancestral character states were reconstructed from cladistic relationships among endemic haplotypes within sampling locations (calculated individually using the branch and bound algorithm of PAUP). Numbers of endemic haplotypes are shown after taxon names. KEKU was excluded from the analysis because it lacked endemic haplotypes.

distantly-vicariated populations. This reasoning is central to PAE, or "parsimony analysis of endemicity," a little-used and little-explored biogeographic technique founded by Rosen and Smith (Rosen, 1984; Rosen, 1985; Rosen and Smith, 1988).

In its original formulation, PAE combined data on species distributions with the cladistic methodology of Hennig (Hennig, 1966) to build cladograms of geographic areas. This technique considers entire geographic areas as "taxa" and species as characters. A species inhabiting areas A and B but not C would count as a synapomorphy supporting a clade linking A and B to the exclusion of C. Parsimony criteria can be applied in the usual sense to find the optimal area cladogram. Taxa may also be autapomorphic or symplesiomorphic, in which case they are uninformative for reconstructing area relationships (Rosen, 1988). Ambiguity in character state assignment (can a taxon ever be proven absent from an area?), character weighting, and the ecological interdependence of taxa are but some of the issues which have not been fully addressed for this method. Although its reasoning is sound, the technique should therefore be considered experimental.

PAE was applied to the chimpanzee genetic data at the allelic level. Sampling locations were defined as taxa and were scored for the presence or absence of each of the 24 identified eurytopic alleles (see Chapter 3). PAUP (Swofford, 1993) was used with the heuristic search option find the most parsimonious tree(s) linking all 19 populations. To estimate statistical confidences associated with observed topological groupings, 1000 heuristic bootstrap replicates were run. Because 634 equally-parsimonious trees were found, a 50% majority-rule bootstrap consensus tree was calculated in which all possible groupings were allowed. The tree was rooted using Lundberg rooting (Lundberg, 1972) and an ancestral population containing no haplotypes. This rooting method is advocated by Rosen and Smith (1988) under the reasoning that it represents "a locality occurring so far back in geological time that none of the taxa in the sample set of localities had yet evolved."

In the present context, it would represent a locality in which none of the alleles in the sample had yet evolved.

Results are presented in Figure 5.9. Bootstrap values on the tree are generally low, with two exceptions. A bootstrap value of 63% supports the monophyly of the two Budongo Forest locations. A bootstrap value of 66% supports a grouping of a Kibale Forest population (KEKU) with a Semliki Forest population (SINI). Budongo Forest monophyly was also observed in two of the three allele frequency distance trees (Figure 5.7) and in the tree of ancestral haplotypes (Figure 5.8). Monophyly of the Budongo Forest locations would be expected for obvious reasons. However, monophyly of locations within other multiply-sampled forests (Semliki and Kibale) is not supported.

The PAE cladogram does not support the hypothesis of population relationship schematized in Figure 5.1. Indeed, the topology presented in Figure 5.1 was a less parsimonious explanation of the data (length = 47) than the reconstructed PAE tree (length = 45) when analyzed using MacClade (Maddison and Maddison, 1992). The PAE tree also lacks long internal branches. This property likely results from the small sample size of characters (24). Because this number is only slightly greater than the number of taxa (19), phylogenetic reconstructions will necessarily be problematic (Swofford, 1993). It is, however, also possible that the pattern observed is real: namely, that the real phylogeny is truly a "star phylogeny" in which resolution is low as a result of the nearly simultaneous genesis of all present-day populations. If so, then area cladograms constructed from other, independent data should show this same pattern.

Fortunately, an independent data set is available for a subset of the populations from which chimpanzees were sampled. Howard (1991) compiled species lists for 12 Ugandan forests using published records, museum collections and extensive ground survey data. Howard recorded the presence or absence of 425 species of forest trees, 329 species of forest birds, 67 species of forest butterflies, and 12 species of forest primates in each of the 12 forests. The data are preliminary in that species absences were not, in most

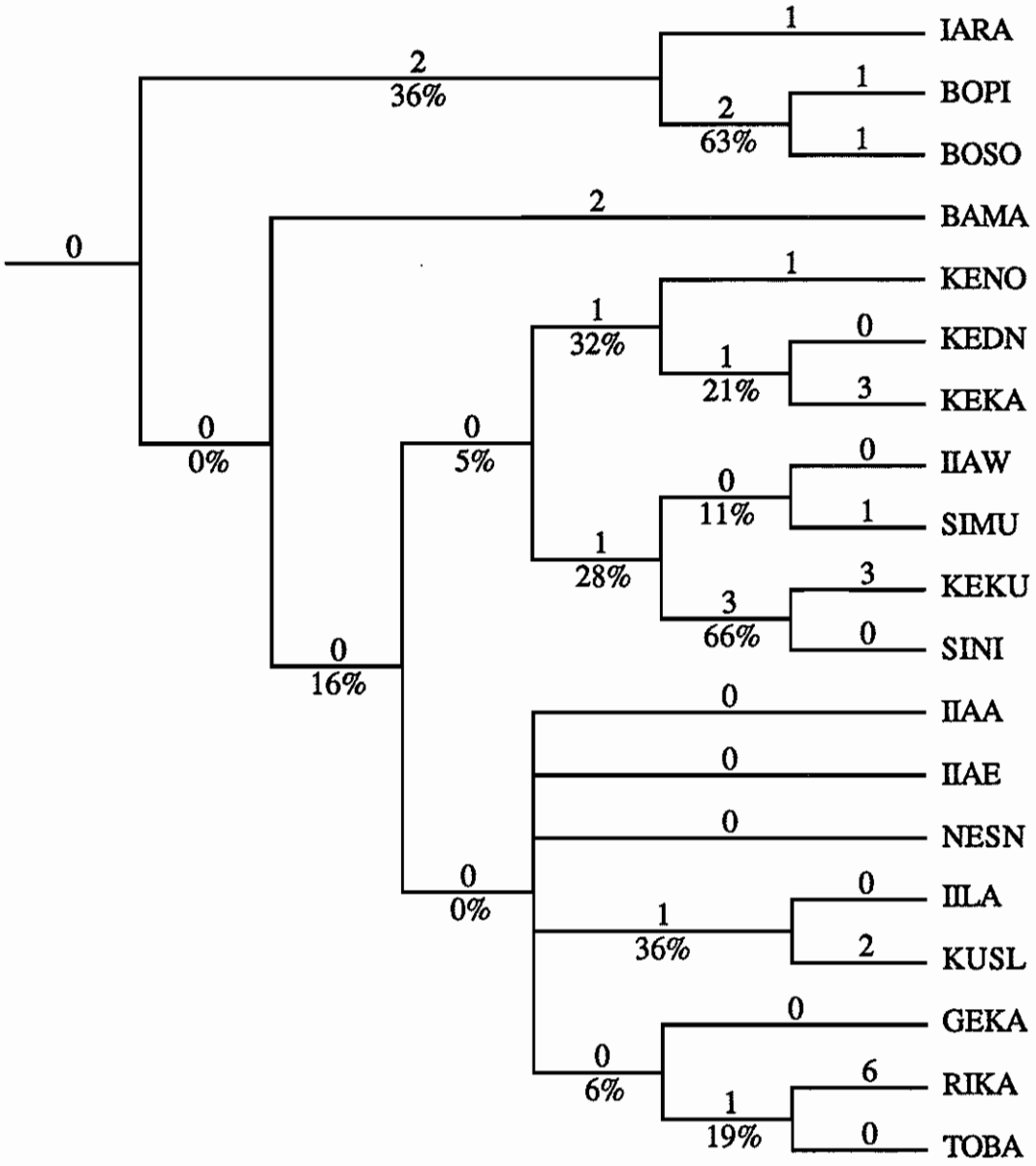


Figure 5.9: Parsimony analysis of endemism using presence/absence of 24 eurytopic haplotypes in 19 sampling locations. The cladogram is a 50% majority-rule consensus tree computed from 1000 bootstrap replicates of the data using the phylogenetic analysis program PAUP (Swofford, 1993) with the heuristic search option. Numbers above branches are minimum numbers of inferred changes. Numbers below branches are bootstrap values. The tree is rooted by the method of Lundberg (1972), using an ancestral population containing no haplotypes. Tree length = 45; rescaled consistency index = 0.207.

cases, absolutely confirmed (Howard, pers. comm.). A more detailed description of the data is available in Howard (1991), and more extensive survey work is currently underway. Nine of Howard's 12 forests contain chimpanzees. Seven of these (Bugoma, Budongo, Itwara, Kibale, Kalinzu, Semliki and Rwenzori) were sampled during the present study.

Parsimony analysis of endemicity was performed on Howard's data using PAUP (Swofford, 1993). The branch and bound search option was used, and 1000 branch and bound bootstrap replicates were run to obtain statistical confidences around observed groupings. Characters were unordered and unweighted. In the case of trees, butterflies and the combined data set, a single most parsimonious tree was found. In the case of birds and primates, more than one equally parsimonious tree was found (2 and 9, respectively). In these latter cases, 50% majority rule consensus trees were calculated. Cladograms were Lundberg rooted (Lundberg, 1972) using a hypothetical ancestral forest containing no taxa (Rosen and Smith, 1988).

Figures 5.10 and 5.11 show the PAE cladograms for trees, birds, butterflies and primates, and for the combined data set of 833 taxa, respectively. The cladograms are similar, with well-defined clades (long internal branch lengths) and high bootstrap values. Bootstrap values for the butterfly and primate cladograms were lower than for the tree and bird cladograms, probably because of relatively low numbers of butterflies and primates. In every case, Rwenzori Forest is most divergent. This would be expected considering Rwenzori Forest's unique ecology (Howard, 1991) and the fact that it was probably the location of a montane Pleistocene forest refuge (Hamilton, 1976; Struhsaker, 1981; Rodgers, Owen and Homewood, 1982; Colyn, 1991). The topology of the cladogram of combined biotas (Figure 5.11) is remarkably robust, having high bootstrap values and long internal branch lengths. It is also generally consistent with the geographic arrangement of forests along the Ugandan forest "archipelago" (see Figure 5.1). Centrally-located forests

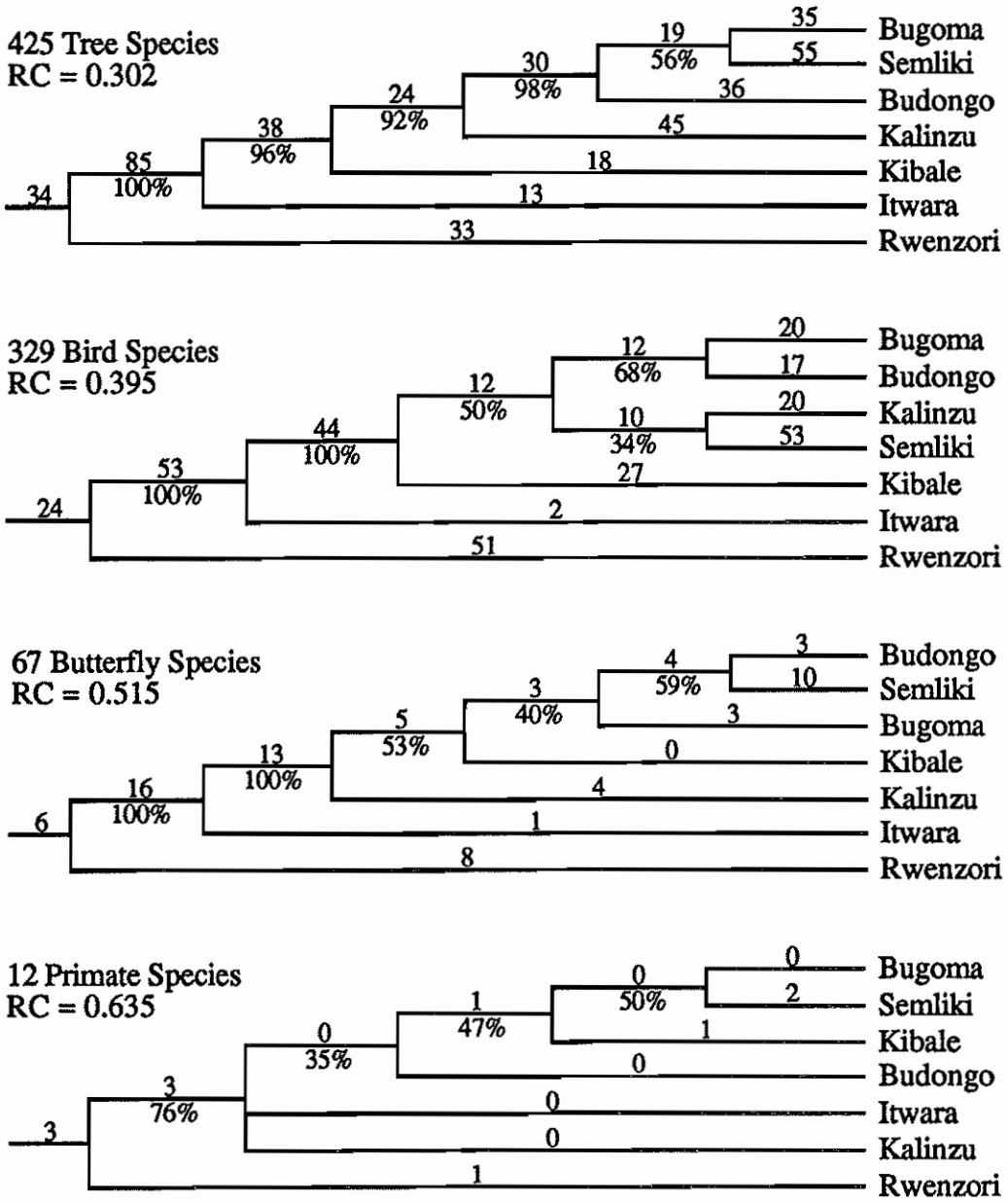


Figure 5.10: Parsimony analyses of endemism for seven Ugandan forests. Data were taken from Howard (1991) and consist of species lists from each forest, compiled from ground survey data. Cladograms are maximum parsimony trees found using the branch and bound algorithm of PAUP (Swofford, 1993) and rooted by the method of Lundberg (1972) with an ancestral location containing no taxa. Numbers above branches are minimum numbers of inferred changes. Numbers below branches are bootstrap values (1000 replicates). Species were equally weighted. Rescaled consistency indices (RC; Farris, 1989) are given for each cladogram.

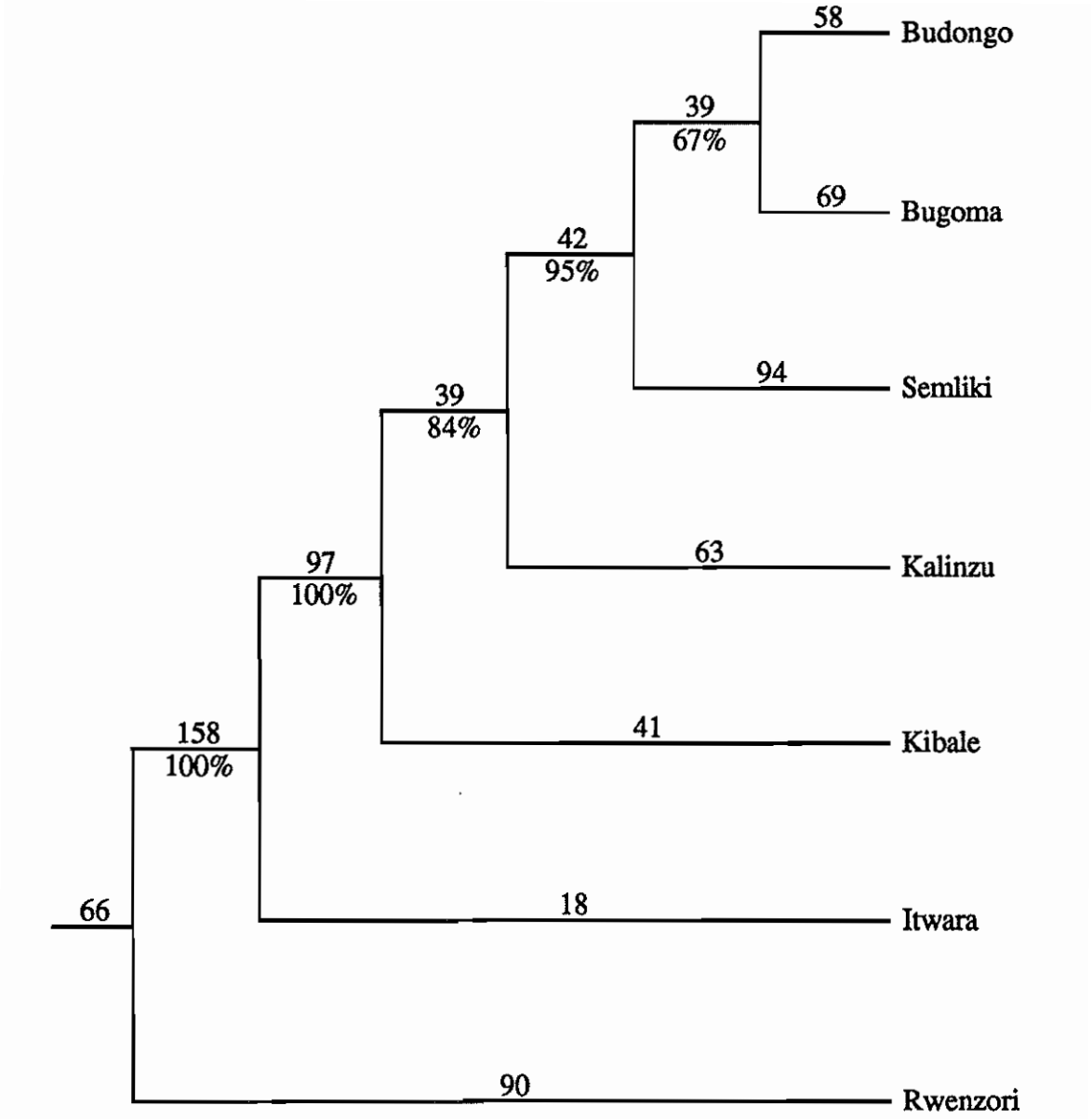


Figure 5.11: Parsimony analysis of endemism for seven Ugandan forests using combined species data for trees (n=425), birds (n=329), butterflies (n=67) and primates (n=12). The cladogram is a maximum-parsimony tree found using the branch and bound algorithm in PAUP (Swofford, 1993), rooted by the Lundberg method (Lundberg, 1973) with an ancestral location containing no taxa. Numbers above branches are minimum numbers of inferred changes. Numbers below branches are bootstrap values (1000 replicates). All characters (species) were equally weighted. Length = 1101; rescaled consistency index = 0.342.

(Itwara, Kibale) branch off after Rwenzori, followed by increasingly more "peripheral" forests (Kalinzu, Bugoma, Budongo). Semliki is an exception in that, although central (it is contiguous with Rwenzori Forest), it clusters with the most peripheral forests (Budongo, Bugoma).

Howard (1991) presents dendrograms of relationships among these forests derived from the same data. Howard's distance-based analysis groups forests according to the overall similarity of their species compositions, which Howard interprets as indicative of ecological similarity among forests within groups. This analysis is thus fundamentally different from the cladistically-derived relationships presented here, which group forests on the basis of inferred shared ancestral characters (species). The topologies generated by the two analyses differ, primarily in the relationships among "ingroup" forests (forests other than Rwenzori); Rwenzori is maximally divergent by both analyses. This observation suggests that the PAE approach may be informative in qualitatively different ways than traditional biogeographic approaches, which are based on overall similarities among species assemblages (Humphries *et. al.*, 1988). In any case, the PAE cladograms of forest taxa bear little resemblance, either topologically or statistically, to the PAE cladogram of eurytopic chimpanzee haplotypes. To the extent that PAE accurately reconstructs history, the forests analyzed appear to be related by a unique, discernible history which is broadly concordant with the predictions of Pleistocene refuge theory. In contrast, chimpanzee populations inhabiting these forests do not conform to this pattern, and do not cluster into well-defined clades by any analysis. This observation may reflect the inadequacy of the genetic data to discerning such relationships. Alternatively, it may indicate that chimpanzee populations are related by a different history, or that relationships among them are truly ill-defined. The latter pattern would suggest either near simultaneity of population origin, or the obscuring of phylogenetic resolution due to extensive gene flow.

Conclusions

"Phylogeographic" analyses of mitochondrial DNA (Avise *et. al.*, 1987; Avise, 1989) have documented intraspecific phylogenetic "gaps" in other species, often corresponding to obvious geographic barriers (Avise, 1994). Such subdivisions typically characterize small, relatively non-dispersive mammals (MacNeil and Strobeck, 1987; Plante, Boag and White, 1989; Riddle and Honeycutt, 1990; Prinsloo and robinson, 1992), birds (Shields and Wilson, 1987; Avise and Nelson, 1989; Edwards and Wilson, 1990; Moore, Graham and Price, 1991; Degnan and Moritz, 1992; Zink and Dittmann, 1993; Seutin *et. al.*, 1994), reptiles and amphibians (Densmore *et. al.*, 1989; Wallis and Arntzen, 1989; Moritz, 1991; Lamb and Avise, 1992), fishes (Bermingham and Avise, 1986; Avise, 1992), and invertebrates (Sounders, Kessler and Avise, 1986; Hale and Singh, 1991; Murray, Stine and Johnson, 1991; Burton and Lee, 1994). However, even large and/or highly-vagile species can exhibit such subdivisioning (Carr *et. al.*, 1986; Baker *et. al.*, 1990b; Cronin, Nelson and Pac, 1991; Wada, Kobayashi and Numachi, 1991). For example, mitochondrial haplotypes of humpback whales (*Megaptera novaeangliae*), a species of high vagility and global distribution, sort into distinct geographic clades due to matrilineal fidelity to migratory routes (Baker *et. al.*, 1990b).

The analyses presented above indicate that *Pan troglodytes schweinfurthii* falls into the other category of taxa which do not show "classical" phylogeographic subdivisioning (e.g. Ball *et. al.*, 1988; Cronin *et. al.*, 1991; Lehman and Wayne, 1991; Arnason, Pálsson and Arason, 1992). Although phylogeographic breaks (representing subspecific divisions) do exist at the *species* level (Morin *et. al.*, 1994a), no such divergent clades exist within the portion of the eastern subspecies sampled in this study. Rather, the eastern chimpanzee pattern is one of semi-restricted gene flow, low overall variability and clinal variation. This pattern most closely recalls that of humans on a global scale (Cann, Stoneking and Wilson, 1987; Merriwether *et. al.*, 1991; Vigilant *et. al.*, 1991). Like humans, chimpanzees are a highly-vagile species, as would be expected from their

documented long-distance ranging ability (Goodall, 1986; Moore, 1992a; Chapman and Wrangham, 1993). Phylogeographic barriers such as major Zaïrian rivers, which demarcate biogeographic subdivisions within and among other primate genera (Colyn, 1991), have apparently not impeded gene flow within eastern chimpanzees to an extent that would create discontinuities on haplotype trees. The lack of haplotype sharing across these rivers (see Chapter 3) implies that they may serve as barriers over time scales too short to have significantly influenced molecular evolutionary processes at the DNA sequence level.

It is therefore not entirely surprising that population trees, in which geographic structure is imposed on the data, had low bootstrap values. All attempts to analyze the genetic data in this way resulted in topologies that were ambiguous, contradictory and of low statistical confidence. This observation supports the conclusions drawn from haplotype tree analyses, and from mismatch distribution analyses presented in Chapter 4, that the genesis of eastern chimpanzee populations was recent and explosive. Any phylogenetic "signal" of the order in which these populations were originally founded (by dispersal or vicariance) has been further swamped by the "noise" of recent migration. This pattern stands in contrast to the inferred history of the forests which these populations occupy. PAE analysis of Howard's (1991) data suggest that relationships among areas are clear, consistent, and statistically well-supported. Furthermore, PAE trees of forest taxa are roughly consistent with the topological predictions of Pleistocene refuge theory outlined at the beginning of this chapter (Figure 5.1).

The possibility does exist that the genetic data are inadequate to the task of detecting population-level evolutionary events of the type predicted by Pleistocene refuge theory. Other, more quickly-evolving loci (e.g. nuclear hypervariable repeat loci) might detect stronger population-level subdivisioning than would mitochondrial DNA, as has been demonstrated for humans (Bowcock *et. al.*, 1994). Genetic data from chimpanzee endosymbionts might also provide enhanced short-term phylogenetic resolution.

However, mitochondrial DNA is, in many respects, optimal for detecting population subdivisioning (Avise *et. al.*, 1987; Avise, 1994). Its pattern of matrilineal transmission would make it especially so for chimpanzees, a species characterized by female dispersal.

To the extent that the mitochondrial data do paint an accurate portrait of chimpanzee phylogeography, chimpanzee genetic evolution has apparently been decoupled from the historical biogeographic processes which have characterized forest evolution within eastern Africa. This observation supports the notion of *P. t.. schweinfurthii* as a highly vagile species, capable of moving across open spaces between forest patches (Kortlandt, 1983; Teleki, 1989; Moore, 1992a). If *P. t.. schweinfurthii* has existed historically in dry, marginal habitats, then a decoupling of forest history from population history would be expected. Perhaps for the same reasons, chimpanzee populations have remained undifferentiated despite biogeographic barriers such as large gaps between eastern forests, or the lacustrine system of lakes Edward and George and the Kasinga Channel. Zaïre's major rivers have apparently acted as effective barriers to short-term migration (Chapter 3), but have not acted so throughout the evolutionary history of the subspecies. Extensive recent gene flow among populations thus makes it difficult to reject the hypothesis that populations in separate refugia were once highly differentiated (Chapman, 1983), and that ancient phylogeographic barriers have only recently been obscured.

Phylogeographic "breaks" may currently exist within the subspecies, but, if so, the populations which would define them were not sampled in this study. Populations at the western extreme of the subspecies range (near the confluence of the Zaïre and Ubangi rivers) are likely candidates, since the major rivers which transect these western forests have been instrumental in the subspecific evolution of Colobine and Cercopithecine primates (Colyn, 1987; Colyn, 1991). Other candidate populations are those in the extreme southwest of the subspecies range (near the headwaters of the Lualaba River), and certain (probably extinct) populations north of the Ubangi River in southern Sudan (Teleki, 1989).

Chapter 6: Genetics and social behavior

Introduction

The application of molecular genetic techniques to the study of primate social behavior is in its infancy. Molecular techniques have the potential to expand current theories of primate social evolution. Specifically, their ability to confirm/reject paternity could elucidate the adaptive forces that have shaped male sociality (Martin, Dixson and Wickings, 1992). Research in the near future will likely continue to focus on two hypotheses: 1) that male dominance predicts male reproductive success, and 2) that genetic relatedness (especially among males) predicts social affiliation. In females, dominance rank is known to predict physiological measures of fertility and observed reproductive output (Fedigan, 1983; Harcourt, 1987; Silk, 1987). In male primates, however, the association between dominance rank and reproductive success has remained elusive, mainly because of the difficulty of ascertaining paternity from morphological clues or behavior. Genealogical relatedness among individuals in wild populations is similarly difficult to ascertain in the absence of continuous multi-generational observational data. Even with genetic data, determining r, the coefficient of relatedness, is infeasible without accurate genealogies with which to calibrate measures of genetic similarity (Gilbert *et. al.*, 1991; Wickings and Dixson, 1992).

The application of molecular genetics to primate behavior has also been limited by the difficulty of obtaining genetic material from free-ranging populations. While recent innovations in non-invasive sampling techniques have opened new methodological doors, these have yet to impact significantly on our understanding of primate sociality in the wild (Höss, Kohn and Pääbo, 1992; Moore, 1992b; Morin and Woodruff, 1992; Taberlet and Bouvet, 1992). In the future, non-invasive techniques will likely be routinely coupled with analyses of "VNTR" (variable number of tandem repeat) loci, now the industry standard

for human DNA fingerprinting (Jeffreys *et. al.*, 1988; Chakraborty and Kidd, 1991). VNTR probes that detect variability in humans have been shown to detect variability in new world monkeys (Gray and Jeffreys, 1989; Dixson *et. al.*, 1992), old world monkeys (Weiss *et. al.*, 1988; Gray and Jeffreys, 1989; Rogers, 1992) and in great apes, including chimpanzees (Ely *et. al.*, 1992; Morin and Woodruff, 1992; Wickings and Dixson, 1992). The ability of these probes to determine parentage and genetic relatedness has led to their popularity in primate behavioral studies (Martin, Dixson and Wickings, 1992).

Some genetic studies have confirmed the expected positive association between reproductive success and male dominance (e.g. Curie-Cohen *et. al.*, 1983; de Ruiter *et. al.*, 1992); others have not (e.g. Duvall, Bernstein and Gordon, 1976; Inoue *et. al.*, 1992). The ambiguity of results to date may be an artifact of the fact that the majority of genetic studies have taken place in captivity (de Ruiter *et. al.*, 1992). Alternatively, male dominance may bear no universal relationship to reproductive success. In chimpanzees (and in primates in general), attaining high status may represent one among several effective strategies, existing in a balance within the community (Tutin, 1980). Other strategies, such as consortships or opportunistic, non-competitive mating, could account for some or even most of the conceptions within a social group. Data from macaques imply that dominance as a reproductive strategy is not effective unless hierarchies are established and stable (Smith, 1981; Inoue *et. al.*, 1990; de Ruiter *et. al.*, 1992). If this phenomenon is generalizable to chimpanzees, then the association between male dominance rank and reproductive success may be stronger in communities in which dominance hierarchies are stable than in communities in which rank positions frequently change.

Specific methods for excluding paternity in chimpanzees using molecular data have already been developed (Ely and Ferrell, 1990; Morin and Woodruff, 1992). Morin *et. al.* (1994b) have applied these methods to the well-studied Kasekela community of chimpanzees at Gombe (Goodall, 1986), providing preliminary data on paternity. Morin

et. al. have also confirmed the hypothesis that genealogical relatedness among males is greater than among females, which is predicted by the male philopatry characteristic of the subspecies. More extensive data and complementary results from other communities will be necessary to show whether the trends described in Gombe are generalizable. Paternity analysis is currently underway at Mahale (Nishida, 1994) and is planned for Budongo (Reynolds, pers. comm.) and Kibale (Wrangham, pers. comm.).

Despite its current popularity, VNTR analysis represents only a subset of the molecular genetic techniques potentially informative to the study of primate behavior. Similarly, paternity exclusion represents only one of the ways in which molecular genetic data can be applied. Analysis of mitochondrial DNA can offer useful, although different, insights. Mitochondrial DNA is maternally inherited. Mitochondrial haplotypes can therefore facilitate "maternity exclusion," since individuals with different mitochondrial haplotypes cannot be mother-offspring pairs. The probability with which maternal siblings share a mitochondrial haplotype is effectively 100%; the probability with which non-siblings share a mitochondrial haplotype is proportional to the haplotype's frequency in the population at large. Therefore, two individuals having different mitochondrial haplotypes have effectively zero probability of being matrilineal siblings. Hypervariable region 1 of the control region (see Chapter 2) is the most quickly-evolving segment of the mitochondrial genome (Kocher and Wilson, 1991). Identity of hypervariable control region sequences is therefore a good indicator of overall haplotype identity between individuals.

This study explores some aspects of chimpanzee social behavior using DNA sequence data from hypervariable region 1 of the mitochondrial control region. Two hypothesis are tested below. First, the hypothesis is tested that matrilineal relatedness predicts social affiliative preference, as assayed by several behavioral measures. This hypothesis is tested specifically for chimpanzee males from Kibale's Kanyawara

community, where detailed data on individual social preferences are available (Wrangham, Clark and Isabirye-Basuta, 1992). Second, the hypothesis is tested that matrilineal relatedness predicts the choice of nesting partners. This hypothesis is explored for 14 communities outside of Kibale for which no direct behavioral data are available.

Social relationships in Kibale Forest's Kanyawara community

Social relationships in *P. t. schweinfurthii* are largely dominated by cooperative associations among adult males. Wherever long-term research on the behavior of eastern chimpanzees has taken place, social relationships among males appear closer and better-defined than relationships among females (Goodall, 1986; Nishida, 1990; Wrangham, Clark and Isabirye-Basuta, 1992). Males are generally dominant to females, and within-male relationships follow a linear dominance hierarchy, notably different from the more tenuous network of social interactions among females (Nishida, 1979; Goodall, 1986; Wrangham, Clark and Isabirye-Basuta, 1992). Differences between male and female sociality are evident in a variety of behavioral indices of social preference, such as association in the same party (Ghiglieri, 1984; Wrangham, Clark and Isabirye-Basuta, 1992), grooming (Nishida, 1990; Wrangham, Clark and Isabirye-Basuta, 1992; Muroyama and Sugiyama, 1994), and nearest neighbor distances (Chapman and White, 1995). These patterns may differ for closely related taxa, although data are scarce. In one population of *P. t. verus*, for example, females feed, groom and travel together frequently (Sugiyama and Koman, 1979; Sugiyama, 1988). In bonobos, close relationships among females and between males and females are the norm (Wrangham, 1986; Muroyama and Sugiyama, 1994; Chapman and White, 1995).

Close affiliative relationships among males function on two levels in eastern chimpanzees: for inter-community aggression and for intra-community politics (Wrangham, 1986). Goodall (Goodall *et. al.*, 1979) was the first to describe inter-community territorial behavior in the subspecies. Intercommunity interactions in eastern

chimpanzees are almost universally hostile, and often involve premeditated, systematic attacks by groups of coalitional males on smaller groups of males from neighboring communities. A chimpanzee community is therefore defined by the spatial and social boundaries at which a core group of coalitional males displays territorial behavior, such as systematic patrolling of community borders (Goodall, 1986). Evidence of territorial behavior has also been reported for Mahale (Nishida, 1979) and Kibale (Holder and Hunt, 1994). By virtue of its ubiquity among the chimpanzee communities thus far studied, and because of its existence in humans, cooperative defense of territory by coalitional males has likely been an important component of chimpanzee behavior at least since the divergence of *Pan* and *Homo* approximately six million years ago (Wrangham, 1987; Ghiglieri, 1989; Wrangham, de Waal and McGrew, 1994).

Within chimpanzee communities, affiliative bonds among adult males can be viewed proximately as political strategies and ultimately as reproductive strategies (de Waal, 1982; Goodall, 1986; Nishida and Hiraiwa-Hasegawa, 1986; Morin, 1993). Cooperative male relationships most likely confer within-community fitness benefits because of the enhanced ability of coalition partners to secure and maintain high dominance rank (de Waal, 1982; Goodall, 1986; Nishida and Hiraiwa-Hasegawa, 1986). The social rank of a male is determined by a complex variety of factors, such as age, physical fitness and intelligence (Goodall, 1986; Nishida and Hiraiwa-Hasegawa, 1986). Nevertheless, chimpanzee males rarely enjoy high rank without the cooperative efforts of one or several partners. All chimpanzee communities in which long-term behavioral observations have taken place contain particularly interactive male-male dyads which engage in relatively high frequencies of cooperative behaviors (Goodall, 1986; Nishida, 1990; Wrangham, Clark and Isabirye-Basuta, 1992). The ecological reasons for alliance formation in primates are better-understood for female-bonded, female philopatric societies (Wrangham, 1980), where alliances function in the cooperative defense of resources. If analogous forces are at

work among male-bonded, male philopatric chimpanzees, then cooperative alliances among males may also confer indirect fitness benefits through increased access to scarce resources.

The most likely scarce resource over which male chimpanzees compete is females. Participation in an alliance relationship would therefore directly enhance fitness, since high rank enhances male mating success (Sugiyama and Koman, 1979; Tutin, 1979; Hasegawa and Hiraiwa-Hasegawa, 1983). Matings in chimpanzees can be divided into three broad categories: opportunistic matings, possessive matings and consortships (Tutin, 1979; Tutin, 1980; Hasegawa and Hiraiwa-Hasegawa, 1983; Hasegawa and Hiraiwa-Hasegawa, 1990; Morin, 1993). Tutin (1979) demonstrated that high rank confers clear reproductive advantages to males during episodes of possessive mating. Tutin observed that, in 69 of 70 possessive matings which were interrupted by another male, the outcome of the interaction was predicted by the interrupter's rank (which determined on the basis of other criteria). Alpha status confers particular advantages during possessive matings, since the alpha male is often the only one capable of monopolizing oestrous females in a group (Sugiyama and Koman, 1979; Tutin, 1979; Hasegawa and Hiraiwa-Hasegawa, 1983). Male rank does not, however, predict success at non-possessive mating attempts (Sugiyama and Koman, 1979; Tutin, 1979; Hasegawa and Hiraiwa-Hasegawa, 1983). This pattern suggests that behavioral dominance may represent only one of several strategies by which male chimpanzees secure mating success. Wrangham (pers. comm.) suggests that male-male alliances in Kibale may function in an even greater capacity than in previously-described populations. In Kanyawara community, dominant coalitionary males coordinate aggression in efforts to mate-guard females cooperatively. The lifetime reproductive benefits of cooperative mate guarding and other such strategies among males will ultimately be elucidated by studies incorporating paternity testing with nuclear DNA markers (Morin *et. al.*, 1994b).

Although the proximate benefits of male coalitionary behavior in chimpanzees are becoming well-understood, the decision rules by which males choose alliance partners are not. Research on captive chimpanzees suggests that alliances are formed and broken on the basis of highly-complex political decisions, in which personality characteristics of the individual community members play central roles (de Waal, 1982). However, captivity may necessarily limit the social choices available to chimpanzees. In the wild, chimpanzees may have access to a much wider range of potential alliance partners. The choice of a partner in the wild probably involves the same sorts of political considerations seen in captivity, but may hinge on other factors as well. These factors may, in turn, explain the greater stability of alliances in the wild than in captivity (Wrangham, pers. comm.).

One plausible hypothesis is that chimpanzees in natural settings form alliances preferentially with kin. Chimpanzees clearly recognize maternally-related kin. Close, lasting bonds are characteristic of mothers and their offspring, and can persist throughout lifetimes (Goodall, 1986). Orphaned infants may be adopted, often by known or suspected maternal siblings (Goodall, 1986). Lack of male investment in infants, coupled with the generally promiscuous mating system of chimpanzees, would make recognition of paternal kin unlikely. However, since primates have been shown experimentally to discern paternal kin through phenotypic matching (Wu *et. al.*, 1980; Fredrickson and Sackett, 1984), patrilineal relatedness should not be discounted as a possible mediator of social affiliation. Nevertheless, an explanation of male social affiliation based on matrilineal relatedness would be predicted by kin selection theory (Hamilton, 1964) and would parallel explanations of female sociality derived for female-bonded primate groups (Wrangham, 1980).

Genetic correlates of social affiliation in Kanyawara

To test the hypothesis that cooperative alliance formation in chimpanzees is predicted by matrilineal relatedness, behavioral data were compiled for 14 chimpanzees

from Kibale's Kanyawara community for which mitochondrial haplotypes were also available (Wrangham, Clark and Isabirye-Basuta, 1992; see Chapter 2 and Appendix 1 for details of sample collection from known individuals). These chimpanzees represent all of the eight adult males in the community (BB, BF, LB, LM, SL, ST, SY, TU), all of the five subadult males (AJ, MS, NJ, RZ, YB) and one adult female (MG). MG was sampled because of her unusually close bond with one of the adult males (TU), suggesting a possible maternal relationship (see below).

Data were compiled from two types of observations: party compositions (PC) and ten-minute focal studies (TMS; see Wrangham, Clark and Basuta, 1992, for a complete description of sampling methods). PC data consisted of 21,396 observations of the identities of individuals comprising a party, made by observers singly or in pairs between 7/88 and 5/93 (Wrangham, Clark and Isabirye-Basuta, 1992). Observations were made every 15 minutes and were therefore not independent. Furthermore, individuals were not represented equally. For example, RZ was killed in an episode of apparent intercommunity aggression in 1992, leaving him necessarily underrepresented in subsequent observations (Holder and Hunt, 1994). TMS data were compiled from observations made between 3/93 and 3/95. Raw data consisted of rotating focal observations of individuals in which grooming interactions and nearest-neighbor distances were recorded as a point sample every two minutes (Wrangham, Clark and Isabirye-Basuta, 1992). These data were edited for the present analysis to exclude all but the fifth (final) observation in each TMS, making each of the resulting 1470 observations relatively independent.

Three indices of social interaction were calculated using PC and TMS data. Dyadic association indices (DAI) were calculated from PC data for all pairs of individuals according to the formula DAI = c/(a+b+c), where a = the number of observations containing individual a without b, b = the number of observations containing individual b without a, and c = the number of observations containing both a and b. Although this

"twice-weight index" is inherently biased towards underestimation of associative preference (Cairns and Schwager, 1987; Ginsberg and Young, 1992), it was chosen because of its frequent use in studies of chimpanzee behavior (Nishida, 1968; Ghiglieri, 1984; Wrangham, Clark and Isabirye-Basuta, 1992). TMS data were used to calculate a "simple ratio" index of grooming preference, GP = g/c, where g = the number of grooming interactions recorded between two individuals and c = the total number of observations in which both individuals were present (Cairns and Schwager, 1987). An analogous index was calculated for nearest-neighbor distances.

Matrices of associative preference among all individuals are presented in Figure 6.1 as UPGMA dendrograms (Sokal and Michener, 1958). The dendrograms are standardized to unit length for purposes of comparison. The closest relationship was assigned a distance of zero and the most distant relationship was assigned a distance of 1; intermediate relationships are proportional to their actual (unadjusted) distances. RZ does not appear in the dendrograms for grooming and nearest neighbor distance because of his death in 1992. To test the consistency of the three measures, correlation tests were performed on the unadjusted matrices according to the method of Hemelrijk (1990a,b). Hemelrijk's Kr test is a modified form of the Mantel test of matrix correlation (Mantel, 1967). It is superior to the Mantel test for studies of social interaction within groups because it considers intra-individual variation; it is also a more powerful test for matrices containing many "ties" (Hemelrijk, 1990a,b). Kr values and one-tailed statistical probabilities were calculated based on 2000 matrix permutations according to the suggestions of Hemelrijk (1990a) using the computer program MATSQUAR, provided by the author.

The three indices of social affiliation were highly correlated. For the correlation between DAI and nearest neighbor distance, Kr = 235, and p = 0.0005. For the correlation between DAI and grooming, Kr = 181 and p = 0.0015. For the correlation between nearest neighbor distances and grooming, Kr = 250 and p = 0.0010. These associations

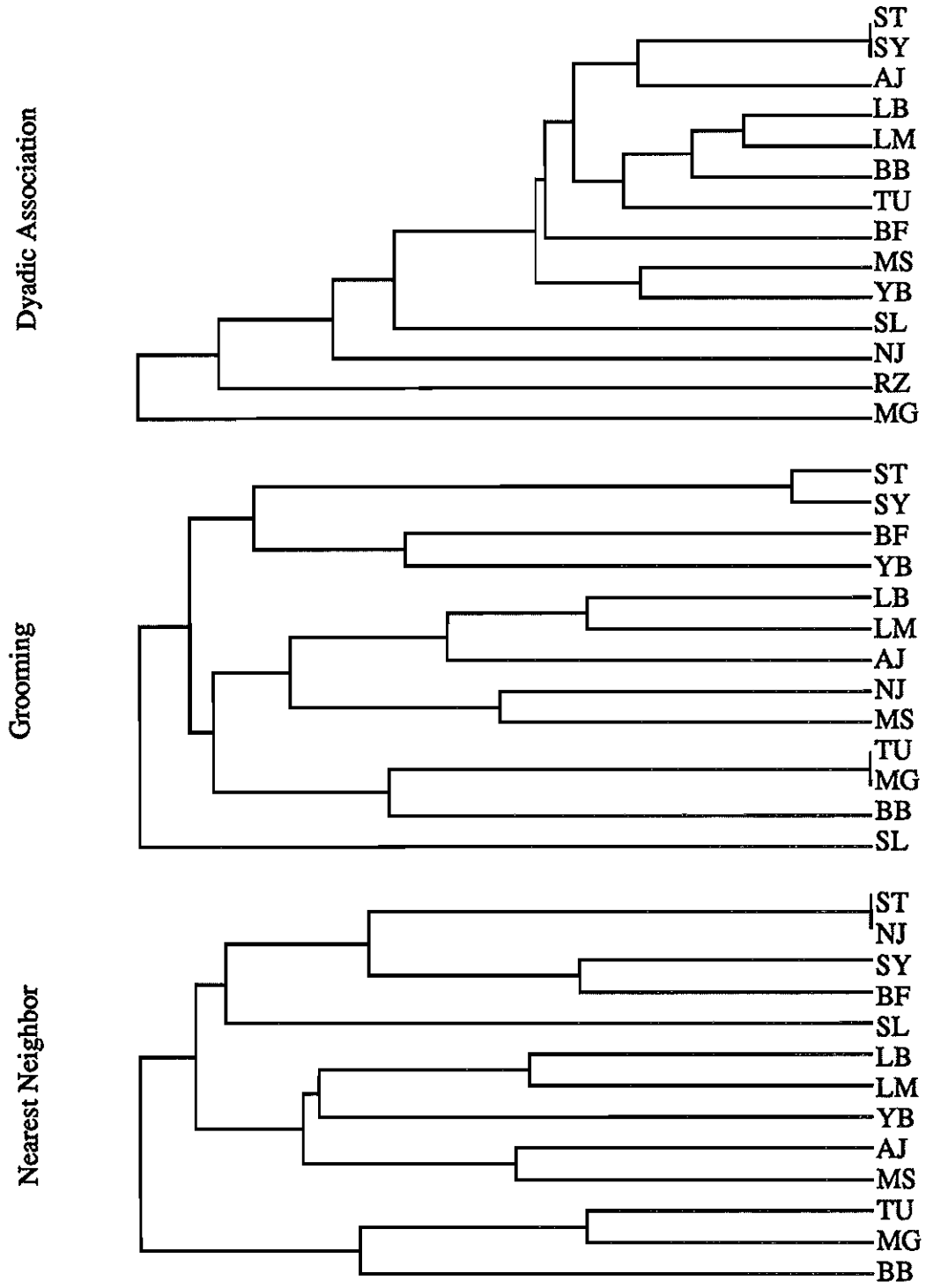


Figure 6.1: UPGMA dendrograms of social affiliation among Kanyawara individuals calculated for three behavioral measures. Dendrograms are standardized to unit length for comparison.

were also tested using standard Mantel tests; probabilities were virtually identical in each case. The concordance among the three measures of social affiliation is encouraging since the measures are experimentally independent. There is no *a priori* reason to expect autocorrelational effects due to interdependence of sampling procedures or biases in the mathematical calculation of associative indices. Rather, correlation among the three measures reflects the fact that affiliative preferences in the chimpanzees studied manifest themselves on different, independent behavioral levels. It is important to note, however, that the dendrograms are not identical, presumably because of a combination of sampling error and real variation in the expression of affiliative preferences across the three behavioral dimensions examined.

Correlation among the three matrices justifies the calculation of a combined matrix of social affiliation, in which each cell represents an average associative index, calculated as the mean standardized distance across all three behavioral measures. This combined dendrogram is presented in standardized form in Figure 6.2. Values for RZ, who appeared only in the DAI matrix, are mean values for grooming and nearest neighbor distances calculated across all other individuals for which data were available. The dendrogram in Figure 6.2 is behaviorally meaningless in that it combines quantitative behavioral data calculated in different units, but it is useful for identifying consistently-affiliative dyads in a qualitative sense.

Four dyads stand out as particularly close. The closest, ST-SY, are the alpha and beta males, respectively, who maintained behavioral dominance in Kanyawara for at least eight years (Wrangham, Clark and Isabirye-Basuta, 1992). Their close relationship is evident in the separate dendrograms for DAI and grooming, but not in the dendrogram for nearest neighbor distance (most likely because small sample sizes for NJ and BF created artifactually close relationships with ST and SY respectively). The next closest relationship, that between TU and MG, is unusual in that MG is an adult female and TU an

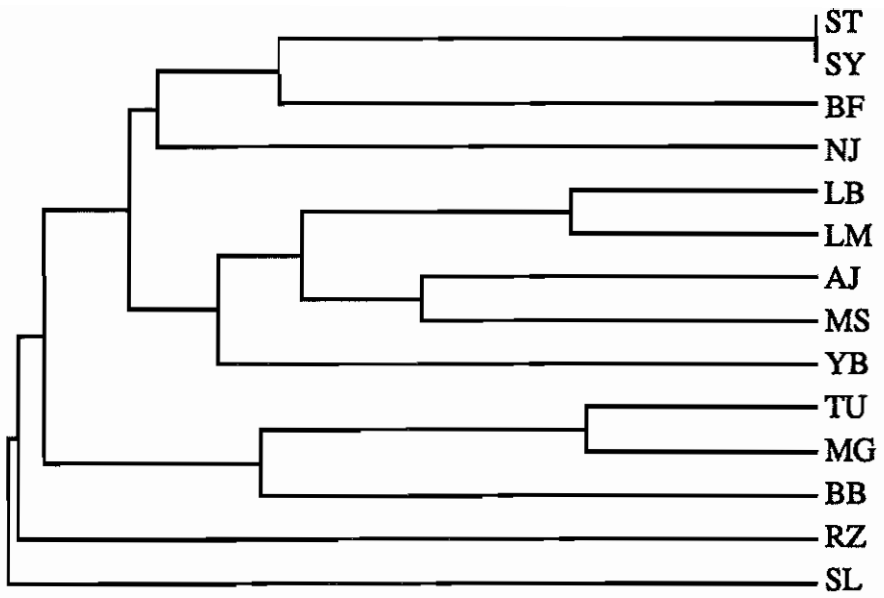


Figure 6.2: Average UPGMA dendrogram of social affiliation among Kanyawara individuals calculated from the means of dyadic association, grooming and nearest neighbor indices. The dendrogram has been standardized to unit length.

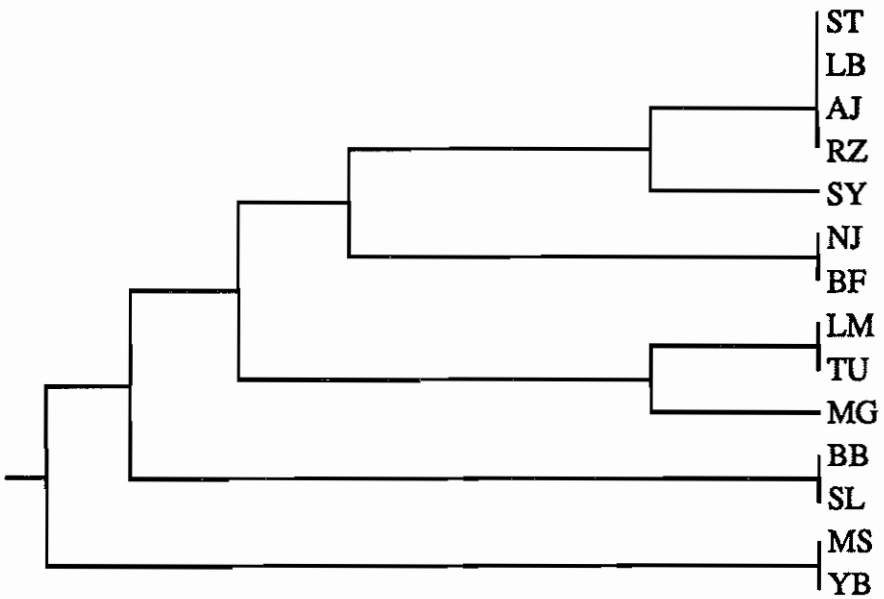


Figure 6.3: UPGMA tree of mitochondrial haplotypes for Kanyawara individuals. Zero branch lengths indicate haplotype identity.

adult male. Wrangham (pers. comm.) qualitatively describes the relationship as asymmetrical, with MG showing intense but generally unreciprocated interest in TU. This relationship is reminiscent of mother-son relationships in Gombe (e.g. Flo and Figan, Melissa and Goblin; Goodall, 1986), suggesting the possibility of a maternal relationship here. The third-closest relationship, LB-LM, represents another affiliative relationship between adult males. It is evident separately in the dendrograms for DAI, grooming and nearest neighbor distance. The fourth-closest relationship, that between the two subadult males AJ and MS, is evident separately only in the dendrogram for nearest neighbor distances. It may represent a nascent cooperative dyadic association between males on the cusp of adulthood. One final relationship should be mentioned. Although BB and TU do not stand out as particularly close in any of the dendrograms presented, this is most likely an artifact of sampling. BB and TU are currently the alpha and beta males, having displaced ST and SY in 1994. Their close affiliative relationship has been documented in detailed social analyses presented by Wrangham, Clark and Basuta (1992).

Figure 6.3 shows a UPGMA tree of genetic relationships for the 14 chimpanzees previously described based on mitochondrial control region sequences. The dendrogram is unstandardized, with zero branch lengths representing actual haplotype identity. For the purposes of discerning potential maternal kin, only two levels of genetic similarity are relevant: identity and non-identity. Mutational events are possible, but their low frequency renders them statistically insignificant to the present analysis (Kocher and Wilson, 1991; Vigilant *et. al.*, 1991; Ward *et. al.*, 1991). Individuals sharing the same haplotype (joined by zero branch lengths) may be members of the same matriline. Individuals with different haplotypes (joined by branches of greater-than-zero length) cannot be members of the same matriline (cannot have shared the same mother).

None of the five aforementioned affiliative dyads share the same mitochondrial haplotype. ST and SY are not maternal brothers, nor are LB and LM, AJ and MS or BB and TU. This observation represents, on its face, a strong rejection of the hypothesis that

Kanyawara males form cooperative alliances on the basis of matrilineal kinship. Surprisingly, MG and TU also do not share the same haplotype and are not therefore mother and son. Extra-genealogical factors must account for this unusual female-male bond.

The possibility still exists, however, that matrilineality predicts affiliative preference *in general* (*i.e.* that close, but non-dyadic, social affiliations are mediated by matrilineal kinship). To test this possibility, matrix correlations were performed between each of the behavioral indices and a matrix of genetic distance based on mitochondrial control region sequences. The behavioral matrices were the unstandardized DAI, grooming and nearest neighbor matrices, calculated as described above. The genetic matrix contained cells with values of either zero or one, representing haplotype non-identity and haplotype identity, respectively. Mantel tests (Mantel, 1967; Smouse, Long and Sokal, 1986) were run on triangular half-matrices (without diagonals) using the computer package "The R Package" (Legendre and Vaudour, 1991). One-tailed probabilities were calculated using a Monte Carlo procedure involving 2000 matrix permutations (Hope, 1968).

The association between nearest neighbor distance and genetic identity was not significant (Mantel $Z = 154$; $p = 0.3378$). Similarly, the association between grooming preference and genetic identity was insignificant (Mantel $Z = 6$; $p = 0.9270$). However, a significant relationship emerged for the comparison between dyadic association indices (DAI) and genetic identity (Mantel $Z = 550$; $p = 0.006$). The same comparisons were also run using Hemelrijk's Kr test (Hemelrijk, 1990a), with the same results (nearest neighbor: $Kr = 1$, $p = 0.4908$; grooming: $Kr = -33$, $p = 0.8821$; DAI: $Kr = 82$, $p = 0.0005$). Mantel tests and Kr tests were also run to compare the standardized combined matrix of social affiliation (represented as a dendrogram in Figure 6.2) with the matrix of genetic identity. The correlation was not significant by either test (Mantel $Z = 4.96$, $p = 0.2650$; $Kr = 23$, $p = 0.2559$).

The analyses described above suggest that, in general, chimpanzees in Kanyawara do not associate preferentially with matrilineal kin. No correlation was observed between genetic relatedness and the two "closest" behavioral measures of social affiliation, grooming and nearest neighbor distance. The significant positive association between haplotype identity and DAI is intriguing, but only indirectly supports the hypothesis that matrilineal kin travel preferentially in the same party. Disregarding mutational events, the probability that two individuals share a haplotype by chance, rather than by common descent from the same female, is proportional to the frequency of the haplotype within the population of distinct matrilines. This probability is impossible to calculate without detailed genealogical information on a population-wide scale. However, it is bound to be high, since the total number of nests sampled in Kibale Forest (56) is represented by only 17 distinct haplotypes (see Chapter 3). Therefore, the positive association between genetics and DAI should be viewed at best as weak support for the hypothesis that chimpanzees associate in the same party preferentially with matrilineal kin. More conservatively, it should be viewed only as a failure to reject the null hypothesis.

The general lack of association between relatedness and social affiliation lends credence to findings from captivity that chimpanzees form bonds primarily on the basis of complex political decisions (de Waal, 1982). This, in turn, speaks to the social complexity of chimpanzee societies under natural conditions, and to the cognitive complexity of individual chimpanzees (van Hoof, 1994). However, it is important to stress that Kanyawara is only one of many chimpanzee communities in East Africa. In light of the great behavioral diversity evident in chimpanzees (Wrangham *et. al.*, 1994), trends from Kanyawara should not be expected to hold for all populations. For example, the trends described above may result simply from the fact that potential alliance partners from the same matriline are scarce in Kanyawara. Kanyawara females are noted for their unusually long interbirth interval (Wrangham *et. al.*, 1996). The probability that a male who is

seeking an alliance partner would be able to select a maternal brother of comparable age and social maturity to himself may therefore be low in comparison to other communities.

However, it is demographically likely that some maternal brother pairs do exist in Kanyawara. A female reproductive lifespan of 15 years (Goodall, 1986; Harvey, Martin and Clutton-Brock, 1987) and an interbirth interval of 6-7 years (Wrangham *et. al.*, 1996) implies that a Kanyawara female may expect three offspring, on average, during her lifetime. There are six unique ordered combinations in which these offspring may appear (M = male; F = female): MMM, FFF, MMF, FFM, MFM, FMF. Thus, 5/9 of males will have an "adjacent" maternal brother, assuming a 50:50 sex ratio. Only adjacent maternal brothers are likely to be close enough in age to serve as potential alliance partners, since the male "window of opportunity" for dominance is probably, at most, ten years in duration (Goodall, 1986). In a community of 13 adult and subadult males (such as Kanyawara), 5/9 X 13, or 7.2 males probably have a maternal brother who is also a potential alliance partner. Therefore, 7.2/2, or 3.6 (between 3 and 4) maternal brother pairs probably exist in Kanyawara. This deduction is supported by the observation that five groups of Kanyawara males are potential maternal brothers in that they share the same mitochondrial haplotype (see Figure 6.3). Given the potential for fraternal cooperative dyads in Kanyawara, it is therefore especially intriguing that they do not occur. In this light, it will be interesting to examine the relationship between matrilineal relatedness and social affiliation in chimpanzee populations displaying a range of demographic and life-history patterns.

Nesting behavior and matrilineal relatedness

Genetic data are most useful when coupled with detailed behavioral observations of individual animals. However, inferences about sociality can sometimes be made even in the absence of direct behavioral observations. Non-observational inferences are facilitated by the fact that chimpanzees leave material traces of their presence scattered across the

landscape (Sept, 1992). Chimpanzees construct arboreal sleeping nests of woven branches both during the day and at night (Nissen, 1931; Goodall, 1962). These nests provide a "fossil record" from which chimpanzee behavior can be inferred (Sept, 1992; Fruth and Hohmann, 1994b). Nests have, for example, been used to estimate population densities (Kano, 1972; Ghiglieri, 1984; Tutin and Fernandez, 1984; Kisubi and Howard, 1990). Nests also represent a dimension of chimpanzee material culture, from which inferences about chimpanzee cultural variability have been made (McGrew, 1985; Fruth and Hohmann, 1994a).

In the present study, chimpanzee served as sources of shed hair from which genetic material was obtained (see Chapter 2). The total number of nests sampled was 431. Some nests did not yield hair, some hair did not yield DNA, and some nests were excluded from the study because of errors during sampling which would have increased the likelihood of contamination by foreign DNA. Therefore, the total number of nests from which genetic data were obtained (229) was considerably smaller than the total number of nests sampled. Of the original 431 nests sampled, data on characteristics of the nests themselves were collected in 357 cases. For these nests, data were recorded on the heights of the nests (in meters) and on the relative ages of the nests (on a 1 to 5 scale, 1 being youngest and 5 being oldest; see Chapter 2). In addition, the number and identities of all nests found in the same "spatial cluster" were recorded. All nests within visual range of each other when observed from the forest floor (regardless of age) were assigned the same single-letter "cluster designation" (see Chapter 2 and Appendix 1). In practice, nests with the same cluster designation formed well-defined "clumps" within forests.

Figure 6.4 shows the frequency distribution of nest heights for the 357 nests from which data were available. The mean nest height (7.83 m, standard error 0.23), the median height (7.0 m) and the range of heights observed (0-27 m) are comparable to values reported for other chimpanzee populations across a range of geographic locations and

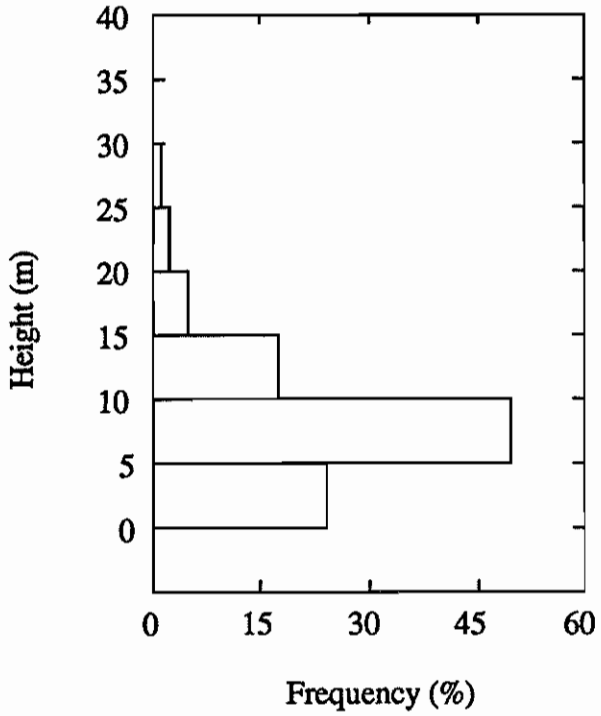


Figure 6.4: Frequency distribution of nest heights for 17 sampling locations (n = 357 observations).

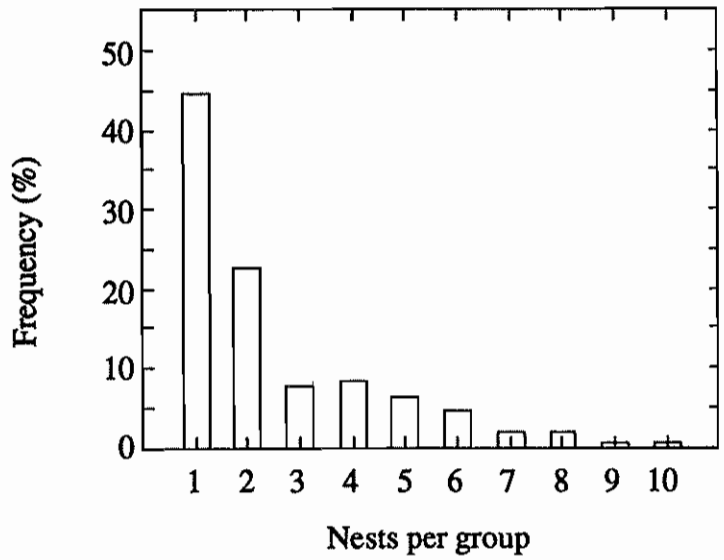


Figure 6.5: Frequency distribution of nest group sizes for 17 sampling locations (n = 153 observations).

habitat types (Fruth and Hohmann, 1994a). Furthermore, the shape of the distribution is comparable to that reported for chimpanzees in Lopé, Gabon (Wroegmann, 1992), as presented by Fruth and Hohmann (1994). These observations suggest that the nests collected in the present study do not represent a severely biased sample of the population of nests at large. Very high nests may be underrepresented in the sample, although the remarkable skills of locally-hired tree climbers (especially the Bambuti of Zaïre) likely minimized this bias.

Figure 6.5 shows the frequency distribution of nest group sizes for the 357 nests (153 groups) for which data were available. Within each location, nests were defined as in the same group if they were assigned the same cluster designation *and* if they were of identical age (measured as described above). For example, a spatial cluster of 10 nests in which half were age two and the other half age three would be scored as two separate groups of five. This age-dependent measure of group identity was chosen to differentiate two sorts of nest clusters: 1) those representing simultaneous nesting events by social groups of chimpanzees and 2) those representing the habitual use of a nesting site by a few individuals over time (see analysis below). Choice of a restrictive definition of group identity may account for the absence of very large groups (≥ 11) in the current study, even though large groups have been reported in low frequency in a number of other populations (Fruth and Hohmann, 1994a). The upper end of the distribution in Figure 6.5 is nevertheless consistent with Wroegman's (1992) data, presented by Fruth and Hohmann (1994), in that it captures the species-typical pattern of frequent single and paired nests. The mean group size (2.44, standard error 0.15) and the median group size (2) were consistent with observations from other study sites (Fruth and Hohmann, 1994a).

Mean nest heights and mean numbers of nests per group are given in Table 6.1 separately for each of the 17 sampling locations for which data were available. Variation is considerable both within and among forests. One-way analyses of variance were used to

Table 6.1: Mean nest heights and mean nest group sizes for 17 chimpanzee populations

	*Location**	*Nests***	*Height†*	*Groups‡*	*Group size†*
1	BAMA	29	7.50 ± 0.67	20	1.45 ± 0.40
2	BOPI	25	10.12 ± 0.72	12	2.08 ± 0.52
3	BOSO	28	5.68 ± 0.69	12	2.33 ± 0.52
4	IARA	33	5.17 ± 0.53	9	3.67 ± 0.60
5	IIAA	36	8.92 ± 0.60	16	2.25 ± 0.45
6	IIAE	22	9.18 ± 0.77	7	3.14 ± 0.68
7	IIAW	16	8.16 ± 0.91	5	3.20 ± 0.80
8	IILA	27	6.06 ± 0.70	10	2.70 ± 0.57
9	KEKA	10	4.65 ± 1.15	3	6.67 ± 1.04
10	KEKU	17	11.59 ± 0.88	7	2.43 ± 0.68
11	KENO	15	11.10 ± 0.94	8	2.00 ± 0.64
12	KESI	7	5.29 ± 1.37	6	2.17 ± 0.74
13	KUSL	31	5.84 ± 0.65	12	2.58 ± 0.52
14	RIKA	18	14.44 ± 0.86	5	3.60 ± 0.81
15	SIMU	13	4.85 ± 1.01	6	2.17 ± 0.74
16	SINI	17	6.21 ± 0.88	7	2.43 ± 0.68
17	TOBA	13	9.23 ± 1.01	8	1.63 ± 0.64

* See chapter 2 for details of sampling locations; Kibale Forest, Sebitoli block (KESI) does not appear in genetic analyses due to small sample size.

** number of nests sampled

† height expressed in meters; errors are standard errors of the mean.

‡ number of nest groups observed

test the hypotheses that nest height and nest group size varied as a function of location. A significant association exists between location and nest height ($r^2 = 0.340$, $p = 0.0001$, 16 degrees of freedom). These results contradict the qualitative claim of Fruth and Hohmann (1994) that no correlation exists between height of nests and location (west vs. central African forests), but are consistent with the conclusions of Baldwin *et. al.* (1981). Variation in nest height likely reflects variation in forest structure, variation in forest ecology, seasonal effects, and cultural differences (Baldwin *et. al.*, 1981; Fruth and Hohmann, 1994a). Similarly, a significant association exists between nest group size and location ($r^2 = 0.201$, $p = 0.01$, 16 degrees of freedom), suggesting that patterns of chimpanzee sociality may vary across populations in analogous ways.

The forces that cause chimpanzees to congregate for sleeping likely hinge on both social and ecological considerations (e.g. mating strategies, protection against predators). If so, nesting in the same group is partly a rough measure of social preference. The rules that govern the choice of sleeping partners may or may not be the same as those that determine other affiliative behaviors. The observation in bonobos that nesting parties are consistently larger than daytime travel parties (Fruth and Hohmann, 1994a) suggests that nighttime may facilitate the expression of centripetal social forces in this species. If generalizable to chimpanzees, this observation suggests that nighttime behavior may reveal hitherto unobserved patterns of sociality. In this light, it is possible that matrilineal kinship, shown in the preceding analysis to be an insignificant social force, might operate in the choice of sleeping partners.

This hypothesis was examined for 14 populations from which both genetic data and data on nest characteristics were available. For each population, a matrix of haplotype identity was created. Pairs of nests yielding the same haplotype were assigned a value of 1; pairs of nests yielding different haplotypes were assigned a value of 0. For these same nests, a matrix of group identity was also created. Pairs of nests from the same spatial

cluster which were also of the same age were assigned a value of 1; pairs of nests from different spatial clusters, or of different ages, were assigned a value of 0. For each population, the genetic identity matrix was correlated with the group identity matrix using a Mantel test (Mantel, 1967). Because of unequal sample sizes among locations, a standardized form of the Mantel Z statistic, r, was calculated (Smouse, Long and Sokal, 1986). This standardized statistic varies between -1 and 1, and is useful for making comparisons across locations; probabilities do not differ from those calculated using the original Z statistic. Probabilities (one-tailed) were calculated from 2000 matrix permutations using a Monte Carlo permutation technique (Hope, 1968).

Table 6.2 presents the results from the 14 matrix correlations. Twelve of the fourteen locations clearly show no association between nesting in the same group and identity of haplotype. Two locations, BAMA and BOSO, show probabilities ≤ 0.05. However, these probabilities (0.049 and 0.023, respectively) are marginal, and most likely represent type 1 error resulting from the large number of independent correlations run (14). A level of significance of 0.0036 ($\alpha' = 0.05/14$) would be required to reject the null hypothesis of no general association, given the number of tests run. The observed values of r for each location also suggest no general trend, in that six are positive and eight negative. Data from fourteen independent locations therefore support the overall conclusion that chimpanzees do not associate in sleeping groups on the basis of matrilineal kinship.

The analysis described above was repeated *without* the criterion that nests need be of the same age to be classified in the same group. Nests were, in other words, considered in the same group solely on the basis of spatial proximity. This analysis is useful since nests decompose at varying rates (Ghiglieri, 1984; Tutin and Fernandez, 1984). A cluster of nests assigned different ages may therefore have been constructed simultaneously. A cluster of nests assigned the same age may analogously have been constructed at different times. Relaxing the age criterion also tests for the possibility that spatial clusters of nests

Table 6.2: Association between nesting in the same group and identity of haplotype for chimpanzees in 14 separate locations*

	*Location***	*n*	*r*†	*p*‡
1	BAMA	13	0.216	0.049
2	BOPI	12	-0.075	0.543
3	BOSO	15	0.291	0.023
4	IARA	12	-0.109	0.513
5	IIAA	13	0.177	0.254
6	IIAE	12	-0.029	0.803
7	IIAW	13	-0.052	0.565
8	IILA	17	0.032	0.864
9	KEKU	14	-0.074	0.472
10	KENO	13	0.074	0.390
11	KUSL	13	-0.047	0.849
12	RIKA	13	-0.086	0.591
13	SIMU	9	-0.112	0.580
14	TOBA	10	0.059	0.894

* Nests were considered in the same "group" only if they were spatially clustered and of the same relative age.

** see Table 2.1 and associated text

† standardized form of the Mantel Z statistic proposed by Smouse, Long & Sokal (1986)

‡ probabilities computed from 2000 matrix permutations following Hope (1968)

were constructed by one or a few individuals during a short period of localized habitat use. This would be the case if individual chimpanzees preferred to re-use nesting sites (Goodall, 1986; Sept, 1992). Relaxing the age criterion is therefore also a test of the hypothesis that double-sampling of individuals was a significant sampling problem (see Chapter 2).

Table 6.3 presents the results of the analysis with the age criterion relaxed. Again, 12 of the 14 populations show no association between genetic identity and nest group identity. The two probabilities ≤ 0.05 (IIAW, $p = 0.038$; TOBA, $p = 0.040$) are marginally significant, and may represent type 1 error. As in the original test, no trend is suggested by values of r, seven of which are positive and seven negative. This observation provides direct evidence that that double-sampling of individuals has not significantly influenced the data.

The overall result from both tests suggests that no association exists between spatial proximity and haplotype identity. This analysis therefore confirms, in fourteen independent populations, that chimpanzees do not nest together on the basis of matrilineal kinship. The rules by which chimpanzees choose to nest together will likely be elucidated by direct behavioral observation.

Conclusions about social behavior

The general lack of association between affiliative preference and matrilineal relatedness demonstrated above is surprising. From long-term field studies, it is known that the closest and most important bond formed in a chimpanzee's life is that with his/her mother (Goodall, 1986; Nishida, 1990). Comparably important in a qualitative sense are the relationships among maternal siblings during childhood, which persist at least up to the time of weaning, and probably through adolescence. Goodall (1986, p. 205) writes, "Bonds that develop between the siblings themselves during these years [juvenile and adolescent years] are likely to endure, particularly those between brothers; this may well be

Table 6.3: Association between nesting in the same spatial cluster and identity of haplotype for chimpanzees from 14 separate locations*

	*Location***	*n*	*r†*	*p‡*
1	BAMA	13	0.135	0.174
2	BOPI	12	-0.075	0.543
3	BOSO	15	0.090	0.316
4	IARA	12	-0.075	0.570
5	IIAA	13	0.130	0.337
6	IIAE	12	-0.024	0.572
7	IIAW	13	0.267	0.038
8	IILA	17	0.032	0.864
9	KEKU	14	-0.052	0.519
10	KENO	13	0.235	0.053
11	KUSL	13	-0.059	0.778
12	RIKA	13	-0.089	0.576
13	SIMU	9	-0.134	0.650
14	TOBA	10	0.401	0.040

* Nests were considered in the same "spatial cluster" even if they were of different relative ages.

**see Table 2.1 and associated text

†standardized form of the Mantel Z statistic proposed by Smouse, Long & Sokal (1986)

‡probabilities computed from 2000 matrix permutations following Hope (1968)

crucial in determining social rank in later life." By all expectation, positive associations between haplotype identity and social affiliation should have emerged in the analyses described above.

The lack of a documented association in Kibale Forest's Kanyawara community could conceivably be an artifact of sampling (only one community was sampled). The positive correlation between DAI and genetic identity also leaves open the possibility that matrilineal relatedness mediates some social decisions, but not others. However, a lack of association between genetic identity and nesting in the same group was documented for 14 communities other than Kanyawara, and suggests lack of a general trend. Nesting in the same group may simply be too crude an index of affiliative preference; it is important to note that mother-offspring relationships are invisible to the analysis of nesting, since mothers and dependent offspring share both sleeping nests and mitochondrial haplotypes. Nevertheless, contrary to the claims of Goodall (1986), the data in general do not confirm the presence of enduring matrilineal bonds.

This solution to this paradox may have to await detailed behavioral observations from a statistically meaningful number of independent chimpanzee communities in East Africa. Nevertheless, the present analysis underscores the need to avoid subjective assessments of genetic relatedness based on unwarranted extrapolations from behavioral data. Goodall's (1986) impression of the importance of close bonds among adult male brothers may have resulted from assumptions about fraternity which were based on behavior. The presumed fraternal relationship between Charlie and Hugh in Gombe's Kahama community and the presumed fraternal relationship between Figan and Faben in Kasekela community are two examples. Without genetic data or multigenerational observation, objective conclusions about genealogy cannot be drawn.

The conclusions drawn above demonstrate the potential utility of mitochondrial genetic data to the study of primate behavior. The central role of matrilineality in theories

of primate social evolution (Wrangham, 1980) underscores the importance of genetic techniques that can reliably reject and confirm maternity. The currently-emerging suite of genetic techniques, coupled with non-invasive sampling, will provide future primatologists with a well-stocked investigative arsenal with which to reconstruct genealogical relationships in general. Such techniques could accelerate the progress of nascent behavioral studies of chimpanzees. Coupled with behavioral data from many populations, they could help discriminate species-typical classes of social relationship from culturally-variable ones.

On a broader scale, comparative genetic studies of separate populations have the potential to reveal hitherto unobservable levels of chimpanzee social organization. The frequency of extra-community conceptions, the extent of long-distance migration and the long-term temporal stability of communities are, for example, unknown in the species. Mitochondrial DNA may prove particularly informative because it traces the movements of females over space and time. Because chimpanzees are a female dispersal species, mitochondrial DNA may be the system of choice for investigating larger-scale behavioral phenomena, such as the transmission of cultural traditions among communities (Wrangham *et. al.*, 1994).

Chapter 7: Conclusions and discussion

Eastern chimpanzee historical biogeography

Two main biogeographic hypotheses were presented in Chapter 1. The first, that eastern chimpanzees have recently experienced a population-size bottleneck, finds strong support in several of the analyses presented in this study. Overall levels of genetic diversity in eastern chimpanzees are low in comparison to other species, including the other hominoids (Chapter 3). Eastern chimpanzee levels are comparable to those in humans, a demographically-similar species known to have experienced a recent bottleneck (Rogers and Jorde, 1995). Mismatch distribution analysis (Chapter 4) provides direct evidence of recent population expansion in the subspecies. The expansion event was considerable in magnitude, from a maximum initial population of 6000 breeding females to a minimum post-expansion population of 70,000 breeding females. The inferred dates of the expansion (between 10.8 and 82.7 kya) are, moreover, consistent with the paleoclimatic prediction that it was driven by the spread of forest habitat during post-glacial climatic amelioration at 12.5 kya (although other climatic events are possible). The expansion was rapid, which also conforms to the predictions of the paleoclimatic model. This conclusion is supported directly by the wave-like shapes the mismatch distributions presented in Chapter 4, and indirectly by the lack of resolution in phylogenetic trees of chimpanzee populations presented in Chapter 5 (implying an explosive, "star phylogeny").

Given these results, it is surprising that no evidence for the second hypothesis presented in Chapter 1, that chimpanzees were restricted to Pleistocene forest refugia prior to expansion, was found. Neither forest area, nor distance from a refuge, nor isolation were good predictors of population genetic diversity. In fact, distance from any single point within the study area failed to explain genetic diversity to any noteworthy degree (Chapter 3). Populations currently inhabiting refuge locations are not significantly more

diverse than are non-refuge populations. IIAA (Zaïrian) and RIKA (Rwenzori Mountain) populations, both of which correspond to the reconstructed locations of Pleistocene forest refugia, do show relatively high genetic diversities. However, other refuge populations (e.g. TOBA, IIAE) ranked consistently low in genetic diversity across all measures (Chapter 3). Similarly there is little genetic evidence that refuge populations were also source populations. Although two of the three haplotype trees presented in Chapter 5 are consistent with a Zaïrian origin for the subspecies, none of the population-level trees examined are. Perhaps the most convincing support for a genetic refuge comes from the analysis of mismatch distribution bimodality presented in Chapter 4. Bimodality in the mismatch distributions of populations in and near to the Rwenzori Mountains suggests a local concentration of ancient chimpanzee lineages. This result, however, must be explained in light of the absence of bimodality for refuge populations in Zaïre.

In short, eastern chimpanzees show convincing genetic evidence of recent population expansion, but poor evidence of restriction to Pleistocene forest refugia prior to expansion. If eastern chimpanzees were restricted to refugia prior to population expansion, then any genetic record of this fact has been obscured by post-expansion gene flow. This combination of results can be most parsimoniously explained by abandoning the notion of *P. t. schweinfurthii* as a forest taxon. As argued in Chapter 1, chimpanzees show a range of behavioral and ecological adaptations to life in a forest habitat (Baldwin *et. al.*, 1981; Goodall, 1986; Wrangham, 1986; Wrangham *et. al.*, 1991; Wrangham *et. al.*, 1993; Fruth and Hohmann, 1994a). The majority of living chimpanzees do indeed inhabit lowland evergreen rainforests (Teleki, 1989). However, chimpanzees also live in many habitats which contain a minimum of closed-canopy forest (Kortlandt, 1983; Teleki, 1989; Wrangham *et. al.*, 1994). These dry habitats consist of small patches of forest (often riparian) interspersed throughout woodland or savannah (McGrew, Baldwin and Tutin, 1981; Moore, 1992a; Sept, 1992). Such habitats have traditionally been considered

"marginal" for chimpanzees (McGrew, Baldwin and Tutin, 1981; Kortlandt, 1983). The results of the present study, however, suggest that ecological conditions characteristic of today's marginal habitats may have typified the eastern chimpanzee environment throughout much of the evolutionary history of the subspecies. If so, then the model of chimpanzee biogeographic history presented in Chapter 1 must be revised.

Under this revised model, tropical evergreen forest would indeed have become restricted to refuge areas during cool, arid phases of the East African climate (Hamilton, 1976; Kingdon, 1981; Struhsaker, 1981; Grubb, 1982; Rodgers, Owen and Homewood, 1982; Colyn, 1991). Outside of refuge areas, a mosaic environment of savannah, woodland and gallery forest would have persisted (Hamilton, 1982; Van Neer, 1984; Maley, 1987; Taylor, 1993). Eastern chimpanzees could have lived in these mosaic environments much as they do today in "marginal" habitats, traveling extensively across open spaces between feeding and nesting sites (McGrew, Baldwin and Tutin, 1981; Sept, 1992). Because population densities in marginal habitats are lower than in forests (Teleki, 1989), overall population numbers of eastern chimpanzees would have been considerably lower than they are today. Re-expansion of forest at 12.5 kya would have led to an increase in the carrying capacity of the environment. This, in turn, would have allowed the population to grow in size. The size expansion, however, would not have involved any concomitant range expansion.

This interpretation would account both for the observed evidence of population expansion, and for the lack of the insular, stepping stone model's predictive power. It is also consistent with the observation that gene flow within the subspecies has been extensive. Chimpanzees in forest environments are extraordinarily vagile, ranging an average of 4.9 km per day for males (3.0 km per day for females), up to a maximum of approximately 10 km per day (Goodall, 1986). At adolescence, females probably disperse over even greater distances (Pusey, 1980). In mosaic habitats, chimpanzees are likely even

more mobile. Although direct behavioral observation has been limited for chimpanzees in dry habitats, home ranges in Ugalla, Tanzania and Mt. Assirik, Senegal have been estimated at 275 km^2 and 560 km^2, respectively (Baldwin, McGrew and Tutin, 1982; Moore, 1992b). These estimates are greater by an order of magnitude than estimates for Gombe (9.6 - 24.0 km^2; Goodall, 1986), and imply that long-distance ranging may be typical of chimpanzees in non-forest habitats.

Subtle restrictions to gene flow (such as those documented by the spatial autocorrelational analyses of Chapter 3 or the tree-based methods of Chapter 5) indicate that localized "pockets" of closely-related haplotypes do exist. However, as demonstrated in Chapter 3, interpopulational migration is extensive (3-5 migrants exchanged among populations per generation, on average) and gene flow has created positive autocorrelational effects over distances of up to 100 km. Biogeographic barriers to dispersal have restricted gene flow only minimally. Eastern chimpanzees show little evidence of phylogeographic partitioning or major population subdivisioning.

This picture of chimpanzee paleodemography is generally consistent with the interpretations of Kortlandt (1983). Kortlandt argues that chimpanzees historically ranged throughout many marginal environments where they are now absent. Recent habitat destruction by humans has preferentially accelerated the extermination of dry forest habitats, which covered a sizable proportion of the species range before the advent of agriculture. Dry forests may have facilitated long-distance migration in the subspecies by linking disparate islands of wetter habitat. Historically, however, chimpanzees may have lived in even patchier, more marginal environments, such as those which likely existed during glacial maxima.

Kortlandt (1983) specifically cites Rabongo Forest in Murchison Falls National Park, Uganda, as a "relict grove," inhabited by five or six marginalized chimpanzees (see Kortlandt, 1983, p. 249). During the present study, Rabongo Forest was visited with the intention of sampling these animals. Despite three days of extensive searching, no

evidence of chimpanzees was discovered. Reports by local park officials suggested that chimpanzees now visit this forest only infrequently. Pabidi Forest Block in Budongo Forest lies approximately 25 km southwest of Rabongo. The intervening grassland contains minimal tree cover, little standing water, and no remaining forest. This area has most likely remained unforested for the past several decades, principally due to frequent human-induced fires. Field workers involved in recent efforts to habituate the Pabidi chimpanzees report that these animals often disappear for extended periods into the dry grasslands to the north.

If the Rabongo chimpanzees are actually migratory animals from Pabidi, then Kortlandt's interpretation of Rabongo as a "relict" population is inaccurate. The Pabidi chimpanzees may habitually migrate across 25 km of dry, open grassland. The ranging habits of chimpanzees in the wetter forests of Gombe, Mahale and Kibale pale in comparison. Such ranging would be more extensive even that that inferred for the relatively isolated population of Bossou, Guinea, separated from its two nearest neighbors by distances of 1 and 10 km (Sugiyama and Koman, 1979). This study suggests that extensive movement across dry environments may have been the norm during periods when East African forests were at a minimum. "Marginal" chimpanzee habitats may thus provide unique insights into components of chimpanzee behavior which, although rare today, may have characterized much of the evolutionary history of the species.

Comparisons with humans

The genetic analyses presented in this study demonstrate many striking similarities between *P. t. schweinfurthii* and *H. sapiens*. Overall levels of genetic variability in the two taxa are similarly low (Chapter 3). In both, the apportionment of diversity follows approximately the same pattern. Despite the fact that human populations occupy a much larger geographic range than do eastern chimpanzee populations, the majority of the total

genetic variation exists within populations in each taxon (Chapter 3). Differences in the apportionment of diversity at other hierarchical levels probably result from the different geographic scales across which eastern chimpanzees and humans can be examined. Eastern chimpanzees cannot be examined on an intercontinental scale, for obvious reasons. The autocorrelational analyses of Chapter 3 demonstrate that humans and eastern chimpanzees resemble one another uncannily at the nucleotide level when examined on similar geographic scales. Phylogenetic trees of chimpanzee haplotypes (Chapter 5) resemble those of humans in showing some geographic clustering, but no major phylogeographic subdivisioning. Chimpanzees also resemble humans in having expanded from a recent population bottleneck (Chapter 4). Moreover, the timing of this event is roughly coincident in the two taxa.

These findings are particularly interesting in light of the multitude of studies which have advocated, on the basis of low human genetic diversity (Ferris, Wilson and Brown, 1981; Wilson *et. al.*, 1985; Kocher and Wilson, 1991), that human evolution has been qualitatively different from that of the other hominoid species (Brown, 1980; Ferris, Wilson and Brown, 1981; Ruvolo *et. al.*, 1994). These studies have been insightful and informative in placing humans in a phylogenetically-appropriate comparative context at the species level. It has been relatively difficult, however, to pursue these comparisons on finer taxonomic levels. This study demonstrates that the suite of genetic characters which has come to support the "mitochondrial Eve" hypothesis of modern human origins turns out not to be uniquely human at all. Humans are, for all intents and purposes, the geographically-expanded genetic equivalent of the eastern chimpanzee subspecies.

At least one genetic similarity between humans and eastern chimpanzees has previously been described. Rogers and Jorde (1995) argue that concordance of the human and eastern chimpanzee mismatch distributions (see Chapter 4) suggests a common cause. They specify the Toba volcanic supereruption at 73.5 kya, principally because of its

geological magnitude and its consistency with their genetically-derived dates for the human population expansion. However, the range of dates of expansion statistically consistent with their data is broad, and includes climatic events of much greater magnitude than the Toba eruption--principally the rapid global deglaciation at 130 kya and the subsequent "switch" of the global climate to warmer, wetter conditions (see Chapter 1 and Figure 1.3). The confidence interval surrounding the reconstructed date of expansion for humans (33 - 150 kya) includes this switch, but is simply too broad to allow the *a posteriori* identification of any particular causal climatic event. Rogers and Jorde present no confidence interval for their eastern chimpanzee data (n = 37; obtained from Morin *et. al.*, 1994a). Results from the present study suggest that the range of acceptable dates for the eastern chimpanzee expansion, while it overlaps considerably with that of humans, is comparably broad (10.8 - 82.7 kya), and excludes the 130 kya climatic shift. In short, evidence for a causal link between the eastern chimpanzee and human expansion events is statistically weak.

Similar "climatic forcing" models have previously been used to explain important human evolutionary events (Vrba, 1985; Vrba, 1989; Vrba, 1993). The origin of bipedal locomotion, for example, is often seen as driven by the gradual replacement of forests by open, savannah-woodland environments (Foley, 1987). One adaptive advantage of bipedalism would thus lie in its energetic efficiency for long-distance travel among habitat patches (Rodman and McHenry, 1980). Since the last common ancestor of humans and chimpanzees was behaviorally and morphologically chimpanzee-like (Wrangham, 1987; Wrangham, de Waal and McGrew, 1994; Pilbeam, 1996), this model of human evolution implies that the transition from (essentially) chimpanzees to australopithecines was driven by the disappearance of forested habitat. Results from the present study, however, suggest that chimpanzees have existed in fragmented, mosaic environments for extended periods of time while still remaining chimpanzees. In fact, the many cyclic forest expansions and contractions which have occurred throughout the Pleistocene (see Chapter 1) have left

chimpanzees more or less unchanged. If the transition from quadrupedal knuckle-walking to bipedalism is to be attributed principally to the widespread disappearance of forest habitat, the present study suggests that the climatic event which caused the transition must have been exceptionally large in both magnitude and duration.

Chimpanzee behavior

Genetic analyses of chimpanzee social behavior are presented and discussed in Chapter 6. As described in that chapter, matrilineal relatedness proves to be a poor predictor of affiliative preference among chimpanzees within a community. This result was true not only for Kibale's Kanyawara community, where detailed behavioral data are available, but also for 14 unstudied communities outside of Kanyawara, where communal nesting was used as an index of social preference. Social interactions among individuals, however, represent only one facet of the chimpanzee behavioral repertoire. The number of chimpanzee communities for which behavioral data are available is steadily increasing, as is the geographic range over which such communities are being studied. It is now apparent that chimpanzees display remarkable cultural diversity (Wrangham *et. al.*, 1994).

Cultural traditions differ among chimpanzee communities separated even by small geographic distances (McGrew, 1992; McGrew, 1994). Although the nature and extent of this variation has been well-documented, few attempts have been made to explain its causes (McGrew, 1992; Wrangham, de Waal and McGrew, 1994). Ecological explanations are plausible only for a subset of chimpanzee behaviors which vary across sites (e.g. nut-smashing, ant-fishing). Other "arbitrary" traditions (e.g. hand-clasp grooming, leaf-grooming) vary in similar ways despite the fact that they appear unrelated to ecological necessity (Wrangham, 1995). Wrangham (1995) describes three non-ecological hypotheses which could explain the observed geographic patterning of chimpanzee cultural traditions.

First, the diffusion of these inventions could be limited (McGrew, 1992). If behaviors are only rarely invented (*i.e.* if convergence is minimal), and if their spread is geographically restricted, then "pockets" of cultural traditions would be expected. Some traditions seem to fit this explanation. Nut-smashing, for example, exists throughout the range of *P. t. verus*, but does not extend eastward beyond the Sassandra River in West Africa (Wrangham, de Waal and McGrew, 1994; Wrangham, 1995). Limited diffusion can not, however, parsimoniously explain the observation that many other traditions, especially arbitrary ones, have disjunct distributions. Hand-clasp grooming, for example, exists in Mahale and Kibale communities, but not in the intervening Gombe community. The limited diffusion hypothesis also implies a slow rate of cultural invention (Wrangham, 1995). This is, however, inconsistent with the observation that individual chimpanzees in captivity and in the wild are exceptionally behaviorally innovative, and that virtually all captive chimpanzee colonies show unique traditions stemming from behaviors invented by captive individuals (de Waal, 1994). Wrangham (1995) therefore proposes that the disjunct distributions of arbitrary traditions may be explained by two alternative mechanisms. First, arbitrary traditions may regularly become locally extinct (transmission failure). The local extinction of traditions could be explained by environmental changes (in the case of ecologically-dependent traditions) or stochastic processes (in the case of arbitrary traditions). Second, arbitrary traditions may regularly evolve independently in different areas (repeat invention).

Because chimpanzees are a female-dispersing species, females are vectors for the inter-community transmission of both cultural traditions and mitochondrial DNA. The geographic patterning of mitochondrial DNA across the landscape therefore provides a framework on which to superimpose the geographic patterning of cultural variation. Concordance between the two patterns would support the limited diffusion hypothesis of cultural evolution. Differences between mitochondrial DNA evolution and the evolution of

cultural traditions are, of course, to be expected. Traditions almost certainly evolve at a faster rate than mitochondrial DNA. Cultural traditions also have a capacity for horizontal transmission which is absent in mitochondrial DNA. Nevertheless, restrictions to the flow of mitochondrial haplotypes across the landscape will almost certainly correspond to cultural "breaks" if diffusion is limited and invention rates are low.

Results from the present study support Wrangham's contention that the limited diffusion hypothesis is insufficient for explaining the geographically disjunct pattern of chimpanzee traditions. Despite marked cultural differences among the chimpanzee communities sampled in this study, no corresponding phylogeographic "breaks" were observed (Chapter 5). As the analyses in Chapter 3 show, all eastern chimpanzee populations tend to share approximately 80-90% of the genetic diversity within the subspecies. Interpopulational differences account for only about 15% of the variation within the subspecies. Furthermore, migration has been extensive and gene flow has effectively "mixed" populations over distances of up to 100 km--distances greater than those which separate communities with different cultural repertoires (e.g. Gombe and Mahale). As the analyses in Chapter 5 show, individual populations are probably related by a "star phylogeny." This observation suggests that the gradual splitting of communities cannot account for the geographic spread of traditions, as implied by a diffusional model of cultural evolution.

The unexpected lack of shared haplotypes between Zaïrian populations separated by large rivers (Chapter 3) suggests that dispersal barriers may operate in some capacity, over short time scales. Such genetically minor barriers may be culturally important and may represent cases in which limited diffusion does, in fact, occur. This hypothesis is supported anecdotally by reports from the indigenous Bambuti people of the Ituri Forest, who claim that a certain heretofore undocumented chimpanzee tradition, removing turtle meat with a stick, occurs on one side of the Epulu River, but may not occur on the other (J.

Hart, pers. comm.). However, temporally-persistent barriers to the diffusion of cultural traditions are unlikely given the absence of such barriers for the diffusion of mitochondrial haplotypes. The diversity of eastern chimpanzee cultures therefore stands in marked contrast to the taxon's genetic homogeneity. This paradox is most parsimoniously explained by concluding that chimpanzee cultural traditions evolve at a much higher rate than do mitochondrial haplotypes, and that traditions routinely become locally extinct.

Chimpanzee conservation

Chimpanzee populations across Africa are disappearing at an alarming rate. Teleki (1989) estimates that chimpanzees are extinct or nearly extinct in nine of the 25 countries throughout which they historically ranged. One of the principal factors causing this decline is the disappearance of chimpanzee habitat (Teleki, 1989). In East Africa, large-scale timber harvesting, mining, and small-scale slash and burn agriculture all contribute to the destruction of the forests in which chimpanzees live (Struhsaker, 1972; Hamilton, 1981; Struhsaker, 1981; Hamilton, 1984; Howard, 1986; Struhsaker, 1987; Weber, 1987; Teleki, 1989; Howard, 1991; Harcourt, 1996). Roads, villages and cultivated land now divide once-continuous chimpanzee habitat (Struhsaker, 1981; Butynski, 1985; Teleki, 1989). The hunting of chimpanzees for meat and for trade on the international pet market further exacerbates this trend (Teleki, 1989). The continuation of such practices despite the implementation of preventative laws suggests that, in the very near future, chimpanzees will be restricted to small populations in a few isolated parks and forest reserves.

The present study demonstrates that eastern chimpanzees have never experienced this kind or degree of population fragmentation at any time during their evolutionary history. Eastern chimpanzee numbers have oscillated cyclically throughout the Pleistocene, but have never declined so quickly or dramatically as at present. Geographic barriers have been relatively unimportant to the subspecies, and gene flow has been extensive across large distances. Never have such daunting barriers as paved roads, human settlements and

agricultural plantations been present. Archaeological evidence suggests that human-induced habitat destruction in East Africa began only 2.5 kya, and has accelerated ever since (Phillipson, 1977; Hamilton, 1984; Hamilton, Taylor and Vogel, 1986; Hamilton, Taylor and Vogel, 1989; Taylor, 1990). Such pressures are far to recent and far too abrupt for chimpanzees to have responded in any adaptive way.

Conservation biologists uniformly acknowledge that habitat destruction and population fragmentation pose serious threats to plant and animal species (Harris, 1984; Saunders, 1991; Woodruff, 1992). Fragmentation enhances the likelihood of extinction for populations within fragments (MacArthur and Wilson, 1967; Terborgh and Winter, 1980; Diamond, 1989; Newmark, 1991). The reasons for this increased vulnerability are largely demographic (Gilpin and Soulé, 1986; Lande and Barrowclough, 1987; Lande, 1988), and include stochastic factors (Lande, 1993), such as local periods of resource depletion (Harris, 1984; Gilpin and Soulé, 1986) and epizootic infections (Dobson and May, 1986; Atkinson, 1989; Packer *et. al.*, 1991). Loss of habitat and habitat fragmentation also reduce the genetic variability of populations (Schonewald-Cox *et. al.*, 1983; Vrijenhoek, 1989; Woodruff, 1989; Woodruff, 1992; Avise, 1994). A wealth of studies have documented alarmingly low levels of genetic diversity in demographically-threatened animal populations (O'Brien *et. al.*, 1983; O'Brien *et. al.*, 1989; McClenaghan, Berger and Truesdale, 1990; Packer *et. al.*, 1991; Sherwin *et. al.*, 1991; Wayne *et. al.*, 1991; Avise, 1994).

Unfortunately, the ultimate consequences of low genetic diversity for population survival are not understood. Many biologists argue that populations low in genetic diversity are prone to extinction through inbreeding depression, the reduced viability of offspring due to increased homozygosity for deleterious alleles (Wright, 1969; Ledig, 1986; Ralls, Harvey and Lyles, 1986). Physiological manifestations of inbreeding such as reduced reproductive potential or reduced infant survivorship are well-documented, and are

certainly important to the management of endangered species in captivity (Ralls, Brugger and Ballou, 1979; Senner, 1980; Ralls and Ballou, 1982; Allendorf and Leary, 1986; Laikre and Ryman, 1991; Packer *et. al.*, 1991). However, it is unclear how such effects might influence the long-term viability of wild populations (Beardmore, 1983; Avise, 1994). All arguments about the supposed importance of genetic diversity to population viability have, to date, been either indirect or speculative. Indeed, one survey of natural butterfly populations was forced to conclude that inbreeding effects do not influence extinction probabilities at all (Ehrlich, 1983). Considering the vastly more tangible and immediate threats of demographic stochasticity, some researchers have advocated that only extra-genetic factors be considered biologically important in analyzing population viability, despite the alluring objectivity of genetic data (Gilpin and Soulé, 1986; Lande and Barrowclough, 1987; Lande, 1988; Lande, 1993).

In the light of this uncertainty, it would be inadvisable to make any conservation recommendations about eastern chimpanzees on the basis of observed levels of genetic diversity in the subspecies or in its individual populations. Populations such as RIKA and IIAA, although relatively diverse in mitochondrial haplotypes, should be accorded no greater conservation value than "depauperate" populations, such as BAMA or IIAE (see Chapter 3). Similarly, it would be inadvisable to deem *P. t. schweinfurthii* more threatened than *P. t. verus* or *P. t. troglodytes* on the basis of its lower genetic diversity (see Chapter 5). Decisions about the allocation of resources to the conservation of specific populations should be based on criteria such as the protected status of the forest, the local attitudes of indigenous people, and the potential of the population to be economically self-sustaining through such ventures as eco-tourism (Struhsaker, 1981; Struhsaker, 1987; Weber, 1987; Teleki, 1989; Harcourt, 1996).

The genetic data collected in the present study are not, however, useless for informing conservation efforts. One tangible result is the observation that all confiscated

chimpanzees in the Entebbe Zoo population belong to the eastern subspecies (see Chapter 5). The natural provenances of these animals are largely unknown, as they were confiscated from secondary or tertiary owners, often *en route* to destinations outside of Africa. Because of the strategic complexity of the international trade in chimpanzees (Teleki, 1989), these animals could have originated from anywhere on the African continent. The demonstration that all 26 Entebbe Zoo animals group phylogenetically with *P. t. schweinfurthii* suggests that limits do exist to the distances over which animals are smuggled within Africa. Nevertheless, the Entebbe Zoo chimpanzees do not tend to cluster together statistically on phylogenetic trees, as do other natural, geographically-defined populations (Chapter 5). Poaching is not, therefore, a localized problem. Rather, the Entebbe Zoo chimpanzees originate from disparate locations across the eastern subspecies range. To halt the illegal capture of wild chimpanzees will therefore require the implementation of stringent international laws which are strictly enforced at the local level.

The Entebbe Zoo population is an unusual resource; few other captive chimpanzee populations likely consist exclusively of a single subspecies. In fact, the majority of captive chimpanzees are probably descendents of wild-caught *P. t. verus* (Teleki, 1989). For this reason, the Entebbe Zoo chimpanzees are attractive candidates for translocation to a wild or semi-wild location. The Entebbe Zoo population could, for example, feasibly be moved to an island, or to a forest where chimpanzees once existed but have become extinct. Mabira Forest in Uganda is one possibility (Howard, 1991). Close to Uganda's capital, this forest could serve as a lucrative eco-tourism attraction if it contained a provisioned chimpanzee population. The genetic consequences of inadvertent gene flow into or out of such a population would be minimal. However, it would be inadvisable to allow any contact, direct or indirect, between the Entebbe Zoo animals and wild chimpanzees, primarily because of the risks of introducing exogenous infectious diseases.

Minimizing the spread of disease is a serious consideration for conservation in general (Dobson and May, 1986; Atkinson, 1989). Epidemic diseases can jeopardize a

population's immediate health and long-term viability (O'Brien *et. al.*, 1983; Packer *et. al.*, 1991). Chimpanzee populations in countries such as Uganda and Tanzania will likely experience increased exposure to exogenous diseases in the near future, as habitat encroachment and eco-tourism force contact with humans. As the 1966 polio-like disease epidemic at Gombe suggests, interspecific disease transmission can have disastrous consequences (Goodall, 1986). It would therefore be useful to determine the ultimate geographic limits over which an introduced disease might spread, and to choose tourism sites which would minimize this process. Results from the present study suggest that, historically, eastern chimpanzees ranged extensively across the landscape and are capable of crossing large open distances. This does not bode well for the efficacy of potential natural epidemiological barriers. However, results from Chapter 3 suggest that large rivers, while not of long-term genetic consequence within the subspecies, may have restricted recent gene flow. In this light, it would be prudent to locate chimpanzee eco-tourism sites at the confluence of large rivers, or at other locations where fluvial or lacustrine barriers would channel, and perhaps limit, the emigration of individuals.

Mitochondrial genetic evolution is slower by several orders of magnitude than the processes of habitat destruction and disease spread discussed above. Mitochondrial DNA therefore yields a picture of the genetic structure of the subspecies which is several thousand years old, at a minimum. It provides, in essence, a "snapshot" of what chimpanzee population structure was like before human intervention. This picture, in turn, provides a yardstick against which to measure present and future changes in the population structure of the subspecies. Habitat fragmentation may, for example, presently be so extensive that chimpanzee populations in the insular forests of Uganda, Tanzania, Burundi and Rwanda have already become completely isolated. Studies of hypervariable nuclear DNA, the DNA of chimpanzee endosymbionts, and, most importantly, direct behavioral observations, have the potential to discover if this is indeed the case. Continued genetic

monitoring of populations could trace such effects through time.

Fortunately, it may be possible to ensure that eastern chimpanzees continue to range as they always have, even in the artificial mosaic of forest fragments, human settlements and agricultural land which will soon dominate their environment. Much attention has been paid to habitat corridors, which channel and facilitate the movement of animals among protected areas (Saunders and Hobbs, 1991; Hobbs, 1992; Simberloff *et. al.*, 1992). Corridors are, however, both difficult to create and costly to maintain. From the standpoint of chimpanzee conservation, corridors may be unnecessary. The present study suggests that chimpanzees are capable of moving across a fragmented, mosaic landscape even in the absence of continuous intervening forest. Chimpanzees often exist in marginal habitats where only small patches of trees are available to provide food and shelter (Kortlandt, 1983). Chimpanzees could be encouraged to move across the modern landscape by ensuring the persistence of very small, relict habitat patches in the intervening open spaces between protected reserves. Chimpanzees routinely nest in habitat patches of only one or two hectares (pers. obs.). The small sizes and economic insignificance of such "stepping stones" would make their maintenance relatively easy. These patches could be created artificially, through tree-planting programs. They could be the focus of educationally-oriented community forestry projects in local villages. The density of important chimpanzee foods could be artificially increased through enrichment planting (Denevan and Padoch, 1988; Dufour, 1990). High densities of figs, for example (Wrangham *et. al.*, 1993), might attract chimpanzees, and could reduce crop-raiding by providing an alternative natural food source.

Conserving chimpanzees requires not only that animals be allowed to persist, but that they persist in ways which preserve as many aspects of their remarkable natural variability as possible. Whatever measures are ultimately adopted to ensure the survival of natural populations of chimpanzees should not, therefore, hinge on any single criterion.

Genetics may provide important information about how chimpanzees have historically used their habitat, or about the geographic ubiquity of illegal poaching. Genetics cannot, however, define the conservation value of a population, or, at present, indicate anything about it's long term survival prospects. Only an intricate combination of political judgement, economic forecasting, ecological modeling and biological inference can inform such decisions. The difficulty of the task is daunting. To ignore it, however, would be unconscionable.

Appendix 1: Catalog of samples used in the study

	Sample[1]	Age[2]	Height[3]	Cluster[4]	Haplotype[5]	Collection Date	Reference/Comments
1	Homo01						Vigilant et. al., 1991
2	Homo02						Vigilant et. al., 1991
3	Ppan01						Horai, 1992
4	Ppan02						Morin et. al., 1994 (Bosandjo)
5	Ptve01						"Hallie" (San Diego Zoo)
6	Ptve02						Morin et. al., 1994 (Ptv5)
7	Ptve03						Morin et. al., 1994 (P_1)
8	Pttr01						(see note 6)
9	Pttr02						(see note 6)
10	Pttr03						(see note 6)
11	BAMA01	3	4.5	A	123	9/27/93	
12	BAMA02	5	3.5	A	44	9/27/93	
13	BAMA03	2	8.5	C	27	9/27/93	
14	BAMA04	3	8.5	D	116	9/27/93	
15	BAMA05	2	8	E	102	9/27/93	
16	BAMA06	3	11.5	G	115	9/27/93	
17	BAMA07	2	14	H	115	9/27/93	
18	BAMA08	2	12	H	115	9/27/93	
19	BAMA09	4	5	K	115	9/28/93	
20	BAMA10	2	4	M	115	9/28/93	
21	BAMA11	2	8	M	115	9/28/93	
22	BAMA12	4	6.5	N	115	9/28/93	
23	BAMA13	4	4	N	115	9/28/93	
24	BOPI01	2	11	A	109	9/18/93	
25	BOPI02	2	8	C	84	9/18/93	
26	BOPI03	2	10	C	40	9/18/93	
27	BOPI04	2	7	C	14	9/18/93	
28	BOPI05	2	10	C	90	9/18/93	
29	BOPI06	2	10	C	40	9/18/93	
30	BOPI07	2	10.5	C	44	9/18/93	
31	BOPI08	2	12.5	D	40	9/18/93	
32	BOPI09	4	8	E	84	9/18/93	
33	BOPI10	3	7	I	82	9/19/93	
34	BOPI11	3	7.5	I	83	9/19/93	
35	BOPI12	3	7.5	I	84	9/19/93	
36	BOSO01	1	7	A	44	9/24/93	
37	BOSO02	1	6	A	44	9/24/93	
38	BOSO03	2	6	B	44	9/24/93	
39	BOSO04	2	4	C	114	9/24/93	

	Sample	Age	Height	Cluster	Haplotype	Collection Date	Reference/Comments
40	BOSO05	3	6	C	39	9/24/93	
41	BOSO06	2	5.5	C	115	9/24/93	
42	BOSO07	2	6.5	C	40	9/24/93	
43	BOSO08	1	4.5	C	82	9/24/93	
44	BOSO09	1	3	C	82	9/24/93	
45	BOSO10	1	3	C	82	9/24/93	
46	BOSO11	3	7	D	99	9/24/93	
47	BOSO12	3	6	D	82	9/24/93	
48	BOSO13	2	4	G	98	9/25/93	
49	BOSO14	2	3	G	38	9/25/93	
50	BOSO15	1	7	H	44	9/25/93	
51	EEZO01				65		Individual ID: Masiko[7]
52	EEZO02				78		Individual ID: Megan[7]
53	EEZO03				86		Individual ID: Sunday[7]
54	EEZO04				75		Individual ID: Robby[7]
55	EEZO05				51		Individual ID: Zakayo[7]
56	EEZO06				10		Individual ID: Catherine[7]
57	EEZO07				34		Individual ID: Kigogoro[7]
58	EEZO08				26		Individual ID: Ruth[7]
59	EEZO09				82		Individual ID: Amina[7]
60	EEZO10				3		Individual ID: Kidogo[7]
61	EEZO11				37		Individual ID: Peace[7]
62	EEZO12				77		Individual ID: Tumbo[7]
63	EEZO13				101		Individual ID: Kate[7]
64	EEZO14				10		Individual ID: Sam[7]
65	EEZO15				107		Individual ID: Connie[7]
66	EEZO16				70		Individual ID: Zesta[7]
67	EEZO17				103		Individual ID: Sophia[7]
68	EEZO18				105		Individual ID: Rufus[7]
69	EEZO19				96		Individual ID: Eddy[7]
70	EEZO20				81		Individual ID: Mika[7]
71	EEZO21				115		Individual ID: Joey[7]
72	EEZO22				115		Individual ID: Becky[7]
73	EEZO23				75		Individual ID: Sally[7]
74	EEZO24				76		Individual ID: Natasha[7]
75	EEZO25				89		Individual ID: Bahati[7]
76	EEZO26				82		Individual ID: Cindy[7]
77	GEKA01				18		Morin et. al., 1994 (Evered)
78	GEKA02				18		Morin et. al., 1994 (California)
79	GEKA03				18		Morin et. al., 1994 (tt3)
80	GEKA04				5		Morin et. al., 1994 (Spindle)
81	GEKA05				4		Morin et. al., 1994 (Gigi)
82	GEKA06				5		Morin et. al., 1994 (Tubi)
83	GEKA07				10		Morin et. al., 1994

	Sample	Age	Height	Cluster	Haplotype	Collection Date	Reference/Comments
84	GEKA08				25		Morin et. al., 1994
85	GEKA09				25		Morin et. al., 1994
86	GEKA10				121		Morin et. al., 1994
87	GEKA11				64		Morin et. al., 1994 (Dh1)
88	GEKA12				32		Morin et. al., 1994 (Frodo)
89	GEKA13				32		Morin et. al., 1994 (Kidevu)
90	GEKA14				31		Morin et. al., 1994 (Wl3)
91	GEKA15				42		Morin et. al., 1994 (Sy2)
92	GEKA16				41		Morin et. al., 1994 (Be1)
93	GEKA17				41		Morin et. al., 1994 (Al2)
94	GEKA18				42		Morin et. al., 1994 (Me1)
95	GEKA19				42		Morin et. al., 1994 (Sandi)
96	IARA01	4	2.5	B	75	9/11/93	
97	IARA02	4	5.5	B	91	9/11/93	
98	IARA03	3	3	B	102	9/11/93	
99	IARA04	3	4.5	B	75	9/11/93	
100	IARA05	1	8	B	40	9/11/93	
101	IARA06	3	2	B	95	9/11/93	
102	IARA07	3	3.5	D	102	9/11/93	
103	IARA08	2	4	E	75	9/12/93	
104	IARA09	2	5	E	114	9/12/93	
105	IARA10	2	8	E	9	9/12/93	
106	IARA11	3	5	F	102	9/12/93	
107	IARA12	3	7	G	74	9/12/93	
108	IIAA01	2	6	E	122	7/28/93	
109	IIAA02	3	7.5	F	2	7/29/93	
110	IIAA03	4	10	G	63	7/30/93	
111	IIAA04	4	11.5	H	1	7/30/93	
112	IIAA05	3	20	I	122	7/31/93	
113	IIAA06	3	8	I	61	7/31/93	
114	IIAA07	2	8	J	122	7/31/93	
115	IIAA08	3	8	K	3	8/9/93	
116	IIAA09	3	5	K	72	8/9/93	
117	IIAA10	1	6	K	11	8/9/93	
118	IIAA11	2	12	L	24	8/10/93	
119	IIAA12	2	8.5	L	24	8/10/93	
120	IIAA13	3	8	L	13	8/10/93	
121	IIAE01	2	4	A	62	8/6/93	
122	IIAE02	2	10	A	71	8/6/93	
123	IIAE03	2	14	A	71	8/6/93	
124	IIAE04	3	7	A	120	8/6/93	
125	IIAE05	2	4	A	62	8/6/93	
126	IIAE06	3	5	A	62	8/6/93	
127	IIAE07	3	10	B	118	8/6/93	

	Sample	Age	Height	Cluster	Haplotype	Collection Date	Reference/Comments
128	IIAE08	3	7	C	62	8/7/93	
129	IIAE09	3	7	C	48	8/7/93	
130	IIAE10	3	6	C	62	8/7/93	
131	IIAE11	3	6	D	62	8/7/93	
132	IIAE12	3	8	D	62	8/7/93	
133	IIAW01	1	5	A	45	8/4/93	
134	IIAW02	1	8	A	88	8/4/93	
135	IIAW03	1	7	A	87	8/4/93	
136	IIAW04	1	7	A	88	8/4/93	
137	IIAW05	1	9	A	58	8/4/93	
138	IIAW06	2	15	A	45	8/4/93	
139	IIAW07	3	18	A	51	8/4/93	
140	IIAW08	2	7.5	A	51	8/4/93	
141	IIAW09	1	5.5	A	51	8/4/93	
142	IIAW10	1	5.5	B	57	8/7/93	
143	IIAW11	2	4	B	66	8/7/93	
144	IIAW12	1	9	B	66	8/7/93	
145	IIAW13	1	9	B	89	8/7/93	
146	IILA01	1	6	A	16	7/21/93	
147	IILA02	1	6	B	15	7/21/93	
148	IILA03	1	8	B	21	7/21/93	
149	IILA04	1	9	B	119	7/21/93	
150	IILA05	2	7.5	C	80	7/21/93	
151	IILA06	2	7.5	C	80	7/21/93	
152	IILA07	3	5.5	D	21	7/21/93	
153	IILA08	1	4	E	110	7/21/93	
154	IILA09	3	7	F	60	7/22/93	
155	IILA10	3	3.5	H	16	7/22/93	
156	IILA11	3	6	H	20	7/22/93	
157	IILA12	1	7	I	23	7/23/93	
158	IILA13	1	3	I	59	7/23/93	
159	IILA14	1	7	I	21	7/23/93	
160	IILA15	1	11	I	101	7/23/93	
161	IILA16	2	3.5	J	16	7/23/93	
162	IILA17	2	6	J	101	7/23/93	
163	KEDN01				113	Jul-93	
164	KEDN02				113	Jul-93	
165	KEDN03				113	Jul-93	
166	KEDN04				113	Jul-93	
167	KEDN05				8	Jul-93	
168	KEDN06				113	Jul-93	
169	KEDN07				113	Jul-93	
170	KEDN08				96	Jul-93	
171	KEDN09				8	Jul-93	

	Sample	Age	Height	Cluster	Haplotype	Collection Date	Reference/Comments
172	KEDN10				22	Jul-93	
173	KEDN11				8	Jul-93	
174	KEDN12				8	Jul-93	
175	KEDN13				8	Jul-93	
176	KEDN14				89	Jul-93	
177	KEKA01	1		A	106	6/24/93	
178	KEKA02	1			97	6/24/93	Individual ID: ST [8]
179	KEKA03	1			96	6/24/93	Individual ID: LM [8]
180	KEKA04	1			97	6/24/93	Individual ID: LB [8]
181	KEKA05	1			116	6/25/93	Individual ID: SY [8]
182	KEKA06	1	3.5	C	96	6/30/93	Individual ID: TU [9]
183	KEKA07	1	7	C	113	6/30/93	Individual ID: NJ [9]
184	KEKA08	1	6	D	75	6/29/93	Individual ID: MS [9]
185	KEKA09	1	5	D	113	6/29/93	Individual ID: BF [9]
186	KEKA10	1	4.5	D	106	6/29/93	Individual ID: BB [9]
187	KEKA11	1			97	Mar-95	Individual ID: AJ [9]
188	KEKA12	1			75	Mar-95	Individual ID: YB [9]
189	KEKA13	1			97	Mar-95	Individual ID: RZ [10]
190	KEKA14	1			106	Aug-95	Individual ID: SL [9]
191	KEKA15	1			94	Aug-95	Individual ID: MG [9]
192	KEKU01	4	12	A	106	7/24/93	
193	KEKU02	5	9	B	17	7/24/93	
194	KEKU03	5	5	C	17	7/24/93	
195	KEKU04	5	7	C	89	7/24/93	
196	KEKU05	5	8	C	96	7/24/93	
197	KEKU06	4	20	A	17	7/26/95	
198	KEKU07	5	7	A	17	7/26/95	
199	KEKU08	5	6	B	17	7/26/95	
200	KEKU09	4	4	B	102	7/26/95	
201	KEKU10	5	12	C	102	7/26/95	
202	KEKU11	5	15	D	36	7/26/95	
203	KEKU12	4	21	A	17	7/26/95	
204	KEKU13	4	16	A	84	7/26/95	
205	KEKU14	5	18	B	51	7/26/95	
206	KENO01	4	10	A	96	7/21/93	
207	KENO02	4	26	B	96	7/21/93	
208	KENO03	5	25	B	96	7/21/93	
209	KENO04	4	8	C	37	7/21/93	
210	KENO05	4	27	D	89	7/21/93	
211	KENO06	4	7	A	97	7/22/93	
212	KENO07	5	15	A	96	7/22/93	
213	KENO08	4	8	A	97	7/22/93	
214	KENO09	4	7	C	12	7/22/93	
215	KENO10	4	13	D	12	7/22/93	

	Sample	Age	Height	Cluster	Haplotype	Collection Date	Reference/Comments
216	KENO11	5	4	C	12	7/22/93	
217	KENO12	5	6	C	89	7/22/93	
218	KENO13	5	3.5	B	96	7/22/93	
219	KUSL01	4	3	A	49	9/5/93	
220	KUSL02	4	8	A	56	9/5/93	
221	KUSL03	2	7.5	A	101	9/5/93	
222	KUSL04	3	4	B	115	9/5/93	
223	KUSL05	4	5.5	B	55	9/5/93	
224	KUSL06	3	2	D	54	9/6/93	
225	KUSL07	3	3	D	113	9/6/93	
226	KUSL08	2	8	F	108	9/6/93	
227	KUSL09	3	7	G	113	9/6/93	
228	KUSL10	3	8.5	G	33	9/6/93	
229	KUSL11	1	4.5	H	52	9/6/93	
230	KUSL12	1	8.5	H	115	9/6/93	
231	KUSL13	1	9	H	53	9/6/93	
232	NESN01				47	1992	(see note 11)
233	NESN02				35	1992	(see note 11)
234	NESN03				112	1992	(see note 11)
235	NESN04				43	1992	(see note 11)
236	NESN05				35	1992	(see note 11)
237	NESN06				46	1992	(see note 11)
238	NESN07				79	1992	(see note 11)
239	NESN08				19	1992	(see note 11)
240	NESN09				112	1992	(see note 11)
241	NESN10				43	1992	(see note 11)
242	NESN11				46	1992	(see note 11)
243	NESN12				46	1992	(see note 11)
244	NESN13				79	1992	(see note 11)
245	RIKA01	1	20	A	97	10/5/93	
246	RIKA02	3	20	A	73	10/6/93	
247	RIKA03	3	17	B	93	10/6/93	
248	RIKA04	3	24	B	7	10/6/93	
249	RIKA05	3	10	B	10	10/6/93	
250	RIKA06	3	10	B	89	10/6/93	
251	RIKA07	3	27	B	6	10/6/93	
252	RIKA08	4	10	C	101	10/7/93	
253	RIKA09	4	12	C	7	10/7/93	
254	RIKA10	4	11	C	30	10/7/93	
255	RIKA11	4	11	C	10	10/7/93	
256	RIKA12	1	11	D	84	10/7/93	
257	RIKA13	1	10	D	34	10/7/93	
258	SIMU01	2	3	A	51	7/7/93	
259	SIMU02	3	7.5	B	30	7/7/93	

	Sample	Age	Height	Cluster	Haplotype	Collection Date	Reference/Comments
260	SIMU03	3	7	B	51	7/7/93	
261	SIMU04	2	3	B	51	7/7/93	
262	SIMU05	2	5	B	51	7/7/93	
263	SIMU06	2	3	B	51	7/7/93	
264	SIMU07	2	3	B	30	7/7/93	
265	SIMU08	2	3.5	B	51	7/7/93	
266	SIMU09	2	3	B	30	7/7/93	
267	SIMU10				51	7/8/93	
268	SIMU11				50	7/8/93	
269	SIMU12				29	7/8/93	
270	SIMU13				28	7/8/93	
271	SINI01				51	8/1/91	
272	SINI02				111	8/1/91	
273	SINI03				111	8/1/91	
274	SINI04				100	8/1/91	
275	SINI05				92	8/1/91	
276	SINI06				51	8/1/91	
277	SINI07	5	6	A	51	7/9/93	
278	SINI08	5	6	C	36	7/10/93	
279	SINI09	5	15	D	17	7/10/93	
280	SINI10	5	4.5	E	102	7/11/93	
281	SINI11				17	7/11/93	
282	TOBA01	1	15	A	69	8/10/93	
283	TOBA02	1	15	A	117	8/10/93	
284	TOBA03	2	6	B	73	8/11/93	
285	TOBA04	3	10	B	68	8/11/93	
286	TOBA05	3	9.5	C	85	8/11/93	
287	TOBA06	4	6	C	85	8/11/93	
288	TOBA07	3	11.5	B	104	8/11/93	
289	TOBA08	3	7	B	67	8/11/93	
290	TOBA09	2	8.5	B	104	8/11/93	
291	TOBA10	3	8	B	104	8/11/93	

Notes

1 Sample names are derived from forest/sampling location names given in Table 2.1.
2 Nests were age ranked on a 5-point relative scale (1 = youngest). See Chapter 2 for details.
3 Height was estimated visually in meters (vertical, from the forest floor).
4 Within each location, the same single-letter code was assigned to nests in the same spatial cluster. Nests were defined as in the same spatial cluster if they were within visual range of each other.
5 Haplotype numbers refer to Appendices 2 and 3.
6 Hair was sent by Phil Morin; samples were collected by Jean Wickings at CIRMF, Gabon.
7 Samples from Entebbe Zoo were collected by Christine Manning.
8 Hair was collected from the forest floor after an observed self-grooming bout.
9 Hair was collected from an individually-identified nest.
10 Hair was collected from the deceased chimpanzee.
11 Samples from Nyungwe Station were collected under the supervision of Elizabeth Williamson.

Appendix 2: DNA sequences used in the study[1]

```
[                  10         20         30         40         50]
[                  .          .          .          .          . ]

Homo01       .G....GG..G.....C........A..C.CC...C.A.............   [50]
Homo02       .G....GG........C........A..C.CC...C.A.............   [50]
Ppan01       ..........G.....C...........C.....C..-T............   [49]
Ppan02       ??????????????????????????????????????T.?...?.....   [50]
Ptve01       ......................C.....C........-............   [49]
Ptve02       ??????????????.........C.....C........-............   [49]
Ptve03       ??????????????????????????....C........-............   [49]
Pttr01       .....................................-............   [49]
Pttr02       .....................................-............   [49]
Pttr03       .....................................-............   [49]
1. (ref.)    AAATTTAAGTACCACCTAAGTATTGGCTTATTCATTA-CAACCGCTATGT   [49]
2.           .....................................-............   [49]
3.           .....................................-............   [49]
4.           ??????????????????????.............-............   [49]
5.           ?????????...........................-..?.........   [49]
6.           .....................................-............   [49]
7.           .....................................-............   [49]
8.           .....................................-............   [49]
9.           .....................................-............   [49]
10.          .....................................-............   [49]
11.          .....................................-............   [49]
12.          .....................................-............   [49]
13.          .....................................-............   [49]
14.          .....................................-............   [49]
15.          .....................................-............   [49]
16.          .....................................-............   [49]
17.          ..................................C..-............   [49]
18.          ?????????.........?.................-............   [49]
19.          .....................................-............   [49]
20.          .....................................-............   [49]
21.          .....................................-............   [49]
22.          .....................................-............   [49]
23.          .....................................-............   [49]
24.          .....................................-............   [49]
25.          ???????.......................C.....-............   [49]
26.          .....................................-............   [49]
27.          .....................................-............   [49]
28.          .....................................-............   [49]
29.          ................................G...-............   [49]
30.          ................................G...-............   [49]
31.          ??????????.........??????.......G...-............   [49]
32.          ?????????..........................G...-............   [49]
33.          ................................G...-............   [49]
34.          ................................G...-............   [49]
35.          .....................................-............   [49]
36.          .....................................-............   [49]
37.          .....................................-............   [49]
38.          .....................................-............   [49]
39.          .....................................-............   [49]
40.          .....................................-............   [49]
41.          ??????????.........?????............-............   [49]
```

```
42.  ????????????????...?????????.....-.....  [49]
43.  .....-.....  [49]
44.  .....-.....  [49]
45.  .....-.....  [49]
46.  .....-.....A.....  [49]
47.  .....-.....  [49]
48.  .....-.....  [49]
49.  .....G.....-..G.....  [49]
50.  .....-.....  [49]
51.  .....-.....  [49]
52.  .....-.....  [49]
53.  .....-.....  [49]
54.  .....-.....  [49]
55.  .....-.....  [49]
56.  .....-.....  [49]
57.  .....-.....  [49]
58.  .....-.....  [49]
59.  .....-.....  [49]
60.  .....-.....  [49]
61.  .....-.....  [49]
62.  .....-.....  [49]
63.  .....-.....  [49]
64.  ????????????????????????????.....-.....?..  [49]
65.  .....-.....  [49]
66.  .....-.....  [49]
67.  .....-.....  [49]
68.  .....-.....  [49]
69.  .....-.....  [49]
70.  .....-.....  [49]
71.  .....-.....  [49]
72.  .....-.....  [49]
73.  .....-.....  [49]
74.  .....-.....  [49]
75.  .....-.....  [49]
76.  .....-.....  [49]
77.  .....-.....  [49]
78.  .....-.....  [49]
79.  .....-.....  [49]
80.  .....C.....-.....  [49]
81.  .....C.....-.....  [49]
82.  .....C.....-.....  [49]
83.  .....-.....  [49]
84.  .....-.....  [49]
85.  .....-.....  [49]
86.  .....-.....  [49]
87.  .....-.....  [49]
88.  .....-.....  [49]
89.  .....-.....  [49]
90.  .....-.....  [49]
91.  .....-.....  [49]
92.  .....-.....  [49]
93.  .....-.....  [49]
94.  .....-.....  [49]
95.  .....-.....  [49]
96.  .....-.....  [49]
97.  .....-.....  [49]
98.  .....-.....  [49]
```

```
99.          ....................................-...........   [49]
100.         ..........T.........................-...........   [49]
101.         ....................................-...........   [49]
102.         ....................................-...........   [49]
103.         ....................................-...........   [49]
104.         ....................................-...........   [49]
105.         ....................................-...........   [49]
106.         ....................................-...........   [49]
107.         ....................................-...........   [49]
108.         ....................................-...........   [49]
109.         ....................................-...........   [49]
110.         ....................................-...........   [49]
111.         ....................................-...........   [49]
112.         ....................................-...........   [49]
113.         ....................................-...........   [49]
114.         ....................................-...........   [49]
115.         ....................................-...........   [49]
116.         ....................................-...........   [49]
117.         ....................................-...........   [49]
118.         ....................................-...........   [49]
119.         ....................................-...........   [49]
120.         ....................................-...........   [49]
121.         ???????.............................?-...........   [49]
122.         ....................................-...........   [49]
123.         ....................................-...........   [49]

[                    60         70         80         90         100]
[                    .          .          .          .          .  ]

Homo01       ..........................................C....AT   [100]
Homo02       ....T..............T.................G....C.C..AT   [100]
Ppan01       ..................................AC.T.............   [99]
Ppan02       ..................................A..T.............   [100]
Ptve01       ..........................................C.......   [99]
Ptve02       ..........................................C.......   [99]
Ptve03       ..........................................C.......   [99]
Pttr01       ................................................   [99]
Pttr02       ................................................   [99]
Pttr03       ................................................   [99]
1. (ref.)    ATTTCGTACATTACTGCCAGCCACCATGAATATTGTACAGTACTATAATC   [99]
2.           ................................................   [99]
3.           ................................................   [99]
4.           ........................?.......................   [99]
5.           ............................?...................   [99]
6.           ................................................   [99]
7.           ................................................   [99]
8.           ................................................   [99]
9.           ................................................   [99]
10.          ................................................   [99]
11.          ................................................   [99]
12.          ................................................   [99]
13.          ................................................   [99]
14.          ................................................   [99]
15.          ................................................   [99]
16.          ................................................   [99]
17.          ................................................   [99]
```

```
18.  ..................................................  [99]
19.  ..................................................  [99]
20.  ..................................................  [99]
21.  ..................................................  [99]
22.  .................................................T  [99]
23.  ................................................C.....T  [99]
24.  .................................................T  [99]
25.  ..................................................  [99]
26.  .................................................T  [99]
27.  ..................................................  [99]
28.  ..................T...............................  [99]
29.  ..................................................  [99]
30.  ..................................................  [99]
31.  ..............................?........??........  [99]
32.  ..................................................  [99]
33.  ..................................................  [99]
34.  ..................................................  [99]
35.  .C................................................  [99]
36.  ..................................................  [99]
37.  ..................................................  [99]
38.  ..................................................  [99]
39.  ..................................................  [99]
40.  ..................................................  [99]
41.  .C..T.............................................  [99]
42.  .C................................................  [99]
43.  ..................................................  [99]
44.  ..................................................  [99]
45.  ..................................................  [99]
46.  ..................T...............................  [99]
47.  ..................................................  [99]
48.  ..C...............................................  [99]
49.  .C..............................................C....  [99]
50.  ..................T...............................  [99]
51.  ..................T...............................  [99]
52.  ..................................................  [99]
53.  ..................................................  [99]
54.  ..................................................  [99]
55.  ..................................................  [99]
56.  ..................................................  [99]
57.  ..................................................  [99]
58.  ..................................................  [99]
59.  ..................T...............................  [99]
60.  ..................T...............................  [99]
61.  ..................................................  [99]
62.  ..................................................  [99]
63.  ..................................................  [99]
64.  ..............................?........??.....?..  [99]
65.  ..................................................  [99]
66.  ................................C.................  [99]
67.  ..................................................  [99]
68.  ..................................................  [99]
69.  ..................................................  [99]
70.  ..............................................C....  [99]
71.  ..................................................  [99]
72.  ..................................................  [99]
73.  ................................A.................  [99]
74.  ..................................................  [99]
75.  ..................................................  [99]
```

```
76.          ..................................................   [99]
77.          ..................................................   [99]
78.          ..................................................   [99]
79.          ..................................................   [99]
80.          ..................................................   [99]
81.          ..................................................   [99]
82.          ..................................................   [99]
83.          ..................................................   [99]
84.          ..................................................   [99]
85.          ..................................................   [99]
86.          ..................................................   [99]
87.          ..................................................   [99]
88.          ..................................................   [99]
89.          ..................................................   [99]
90.          ..................................................   [99]
91.          ..................................................   [99]
92.          ..................................................   [99]
93.          ..................T...............................   [99]
94.          ................T.................................   [99]
95.          ..................................................   [99]
96.          ..................................................   [99]
97.          ..................................................   [99]
98.          ..................................................   [99]
99.          ..................................................   [99]
100.         ..................................................   [99]
101.         ..................................................   [99]
102.         ..................................................   [99]
103.         ..................................................   [99]
104.         ..................................................   [99]
105.         ..............................................C...   [99]
106.         ..............................................C...   [99]
107.         .C............................................C...   [99]
108.         ..............................................C...   [99]
109.         ..............................................C...   [99]
110.         ..................................................   [99]
111.         ..C...............................................   [99]
112.         ..................................................   [99]
113.         ..................................................   [99]
114.         ..................................................   [99]
115.         ..................................................   [99]
116.         ..................................................   [99]
117.         ..................................................   [99]
118.         ..................................................   [99]
119.         ..................................................   [99]
120.         ..................................................   [99]
121.         .......?..........................................?..   [99]
122.         ..................................................   [99]
123.         ..................................................   [99]

[                    110         120         130         140         150]
[                    .           .           .           .           .  ]

Homo01       ...TG..C....G..GT....A...AC...AT..-....C.A....CT.C   [149]
Homo02       ...TG..C....G..GT....A...AC...AT..-....C.A....C..T   [149]
Ppan01       .T.T...C............A...AC.T..AT........A....C..C   [149]
Ppan02       ...T...C............A...AC.T..AT........A...?.TT.   [150]
```

```
Ptve01       ..C....C.......G.....A...-.....-.T.......A.....T..   [147]
Ptve02       ..C.................A...-.....-T.....CC.A.....T..   [147]
Ptve03       ..C.................A...-.....-T.....CC.A.....T..   [147]
Pttr01       ..............G.....T...-C....-T........C...G...C   [147]
Pttr02       ....................T...-.....-T........C..TG...C   [147]
Pttr03       ................T....T...-C....-T........C...G...C   [147]
1. (ref.)    ACTCAACTACCTATAACACATCAAA-TCCAC-CCCACATTATAACCTCCA   [147]
2.           ........................-.....-.................   [147]
3.           ........................-.....-.................   [147]
4.           .T...?.........?T........-C....-.....?.C.C.......   [147]
5.           .T.............T........-C....-.......C.C.......   [147]
6.           .T.............T........-CT...-.......C.C.......   [147]
7.           .T.............T........-CT...-.......C.C.......   [147]
8.           .T.............T........-CT...-.......C.C.......   [147]
9.           .T.............T........-CT...-.......C.C.......   [147]
10.          .T.............T........-CT...-.......C.C.......   [147]
11.          ................T........-.....-...G...........T.   [147]
12.          ................T........-.....-...G.............   [147]
13.          ................T........-.....-...G.............   [147]
14.          ................T........-C....-...G.............   [147]
15.          ................T........-C....-...G.....C.G.....   [147]
16.          ................T........-C....-...G.....C.......   [147]
17.          .T.....C........T........-CT...-.........C.......   [147]
18.          .T.............T........-C....-.........C.......   [147]
19.          ................T........-C....-.........C.G.....   [147]
20.          ................T........-C....-.........C.G.....   [147]
21.          ................T........-C....-.........C.G.....   [147]
22.          ............C..T........-CT...-.........C.G.....   [147]
23.          ................T........-C....-.........C.G.....   [147]
24.          ................T........-C....-.........C.G.....   [147]
25.          ................T........-C....-.........C.G.....   [147]
26.          ................T........-C....-.........C.G..C...   [147]
27.          ................T........-C....-.........C.G..C...   [147]
28.          ................T........-C.T..-................T.   [147]
29.          ................T........-C.T..-.................   [147]
30.          ................T........-C.T..-.................   [147]
31.          ................T........-C....-..........?......   [147]
32.          ................T........-C....-.................   [147]
33.          ................T........-C....-.................   [147]
34.          ................T........-C....-.................   [147]
35.          ................T........-C....-.................   [147]
36.          ................T........-C....-.................   [147]
37.          ................T........-C....-.................   [147]
38.          ................T........-C....-.................   [147]
39.          ................T........-C....-.................   [147]
40.          ................T........-C....-.................   [147]
41.          ................T........-CT...-.........C?.......   [147]
42.          ................T........-CT...-.........C........   [147]
43.          ................T........-CT...-.........C........   [147]
44.          ................T........-CT...-.........C........   [147]
45.          ................T....T...-C....-...G.....C.......C   [147]
46.          ................T........-C....-.........C......T.   [147]
47.          ................T........-C....-.........C......T.   [147]
48.          ................T........-.....-.........C........   [147]
49.          ................T........-C....-.................   [147]
50.          ................T........-C....-................T.   [147]
51.          ................T........-C....-................T.   [147]
```

```
52.  ................T........-C....-................T.  [147]
53.  ................T........-C....-................T.  [147]
54.  ................T........-C....-................T.  [147]
55.  ................T........-C....-.......C........T.  [147]
56.  ................T........-C....-................T.  [147]
57.  ................T........-C....-..........G....T.  [147]
58.  ................T........-C....-................T.  [147]
59.  ................T........-C....-................T.  [147]
60.  ................T........-C....-................T.  [147]
61.  ................T........-CT...-................T.  [147]
62.  ................T........-C....-................T.  [147]
63.  ................T........-C....-................T.  [147]
64.  ................T........-C....-..........?.....T.  [147]
65.  ................T........-C....-................T.  [147]
66.  ................T........-C.T..-.........C.G.....  [147]
67.  ................T........-C...T-................T.  [147]
68.  ........................-C....-.........C........  [147]
69.  ................T........-C....-.........C........  [147]
70.  ................T........-C....-.........C......T.  [147]
71.  ................T........-C....-.........C......T.  [147]
72.  ................T........-C....-.........C......T.  [147]
73.  ................T........-.....-.........C......T.  [147]
74.  ................T........-C....-.........C......T.  [147]
75.  ................T........-C....-.........C......T.  [147]
76.  ................T........-C....-.........C........  [147]
77.  ................T........-C....-.........C......T.  [147]
78.  ................T........-C....-.........C........  [147]
79.  ................T........-C....-.........C......T.  [147]
80.  ................T........-C....-.........C......T.  [147]
81.  ................T........-C....-.........C......T.  [147]
82.  ................T........-C....-.........C......T.  [147]
83.  ................T........-C....-.........C......T.  [147]
84.  ................T........-C....-.........C......T.  [147]
85.  ................T........-CT...-.....................  [147]
86.  ................T........-C....-.........C........  [147]
87.  ................T........-C....-.........C........  [147]
88.  ................T........-C....-.........C........  [147]
89.  ................T........-C....-.........C........  [147]
90.  ................T........-C....-.........C......T.  [147]
91.  ................T........-C....-.........C........  [147]
92.  ................T........-C....-.........C........  [147]
93.  ................T........-C....-.........C........  [147]
94.  ................T........-C....-.........C........  [147]
95.  ................T........-C....-.........C........  [147]
96.  ................T........-C....-.........C........  [147]
97.  ................T........-C....-.....................  [147]
98.  ................T........-C....-.........C........  [147]
99.  ................T........-C....-.........C........  [147]
100. ................T........-C....-.........C........  [147]
101. ................T........-C....-.........C........  [147]
102. ................T........-C....-.........C........  [147]
103. ................T........-.....-T.....................  [147]
104. ................T........-C....-................T.  [147]
105. ................T........-C....-.....................  [147]
106. ................T........-C....-.....................  [147]
107. ................T........-C....-.....................  [147]
108. ................T........-C....-.....................  [147]
109. ................T........-C....-.....................  [147]
```

```
110.          ................T........-C....-.................   [147]
111.          ................T........-C....-.................   [147]
112.          ................T........-C....-.................   [147]
113.          ................T........-C....-.......C.........   [147]
114.          ................T........-C....-.................   [147]
115.          ................T........-C....-.................   [147]
116.          ................T........-C....-.................   [147]
117.          ................T........-C....-.........C........   [147]
118.          ................T........-C....-.........C........   [147]
119.          ................T........-C....-.........C........   [147]
120.          ................T........-C....-.........C........   [147]
121.          ................T?.......-C....-.................?   [147]
122.          ................T........-C....-.................   [147]
123.          ................T........-C....-.................   [147]


[                    160         170         180         190         200]
[                    .           .           .           .           .  ]

Homo01        ................A.T...GT..............A...T....C   [199]
Homo02        ................A.....G.........T......A........C   [199]
Ppan01        ........A..T.......A....T....G...TC........GA.....   [199]
Ppan02        ...................A....T........TC........A.....   [200]
Ptve01        ..................................C.....A..GA.....   [197]
Ptve02        .................................TC........GA.....   [197]
Ptve03        .................................TC........GA.....   [197]
Pttr01        ...................T..............................   [197]
Pttr02        ....CC.............T..............................   [197]
Pttr03        ................T.T..............................   [197]
1. (ref.)     CCCCATGCTTACAAGCACGCACAACAATCAACCCTCAACTGTCACACATA   [197]
2.            ..................................................   [197]
3.            ..................................................   [197]
4.            ...................................?.........?..   [197]
5.            ..................................................   [197]
6.            ..................................................   [197]
7.            ..................................................   [197]
8.            ................T.................................   [197]
9.            ..................................................   [197]
10.           ..................................................   [197]
11.           ..................................................   [197]
12.           ................T.T................................   [197]
13.           ................T.................................   [197]
14.           ................T.................................   [197]
15.           ..................................................   [197]
16.           ..................................................   [197]
17.           ..................................................   [197]
18.           ..................................................   [197]
19.           .................................T................   [197]
20.           .................................T................   [197]
21.           .................................T................   [197]
22.           ..................................................   [197]
23.           ..................................................   [197]
24.           ..................................................   [197]
25.           ..................................................   [197]
26.           ..................................................   [197]
27.           ..................................................   [197]
28.           ..................................................   [197]
```

```
29.  ..................................................  [197]
30.  ..................................................  [197]
31.  .............?.........?..........................  [197]
32.  ..................................................  [197]
33.  ....................................G.............  [197]
34.  ..................................................  [197]
35.  ................T.................................  [197]
36.  ...T............................................... [197]
37.  ...T............................................... [197]
38.  ..................................................  [197]
39.  ..................................................  [197]
40.  ..................................................  [197]
41.  .............?....................................  [197]
42.  ..................................................  [197]
43.  ..................................................  [197]
44.  ..................................................  [197]
45.  ............................G.....................  [197]
46.  ..................................................  [197]
47.  ................T.................................  [197]
48.  ...T............T.T................................  [197]
49.  ..................................................  [197]
50.  ...T.............................................. [197]
51.  ...T.............................................. [197]
52.  ..................................................  [197]
53.  ..................................................  [197]
54.  ..................................................  [197]
55.  ..................................................  [197]
56.  ...T.............................................. [197]
57.  ..................................................  [197]
58.  ..................................................  [197]
59.  ..................................................  [197]
60.  ..................................................  [197]
61.  ..................................................  [197]
62.  ..................................................  [197]
63.  ..................................................  [197]
64.  ..........?..?.........?..................?.......  [197]
65.  ................T.T.................................  [197]
66.  ...T.............................................. [197]
67.  ...T.....................T.....................  [197]
68.  ..................................................  [197]
69.  ..................................................  [197]
70.  ..................................................  [197]
71.  ..................................................  [197]
72.  ................T.................................  [197]
73.  ..................................................  [197]
74.  ................T.................................  [197]
75.  ................T.................................  [197]
76.  ..................................................  [197]
77.  ..................................................  [197]
78.  ..................................................  [197]
79.  ................T.................................  [197]
80.  ................T.................................  [197]
81.  ................T.............C...................  [197]
82.  ................T.................................  [197]
83.  ................T.............C...................  [197]
84.  ................T.............C...................  [197]
85.  .T.....................T.....................  [197]
86.  ..................................................  [197]
```

```
87.          ...T.....................................T.......    [197]
88.          ...T.....................................T.......    [197]
89.          ...T.....................................T.......    [197]
90.          ...T.....................................T.......    [197]
91.          ...T.........................................    [197]
92.          ...T.........................................    [197]
93.          ...T.........................................    [197]
94.          ...T.........................................    [197]
95.          ...T.........................................    [197]
96.          ...T.........................................    [197]
97.          ...T.........................................    [197]
98.          ...T.........................................    [197]
99.          ...T.........................................    [197]
100.         ...T.......................T.....................    [197]
101.         ...T.........................................    [197]
102.         ...T.......................T.....................    [197]
103.         ...T.........................................    [197]
104.         ...T.........................................    [197]
105.         .............................................    [197]
106.         .............................................    [197]
107.         .............................................    [197]
108.         .............................................    [197]
109.         .............................................    [197]
110.         ...T.........................................    [197]
111.         .............................................    [197]
112.         ............................T.....................    [197]
113.         .............................................    [197]
114.         .............................................    [197]
115.         .............................................    [197]
116.         ...T.........................................    [197]
117.         .............................................G....    [197]
118.         .............................................    [197]
119.         ...T.........................................    [197]
120.         ...T.........................................    [197]
121.         ...T.......................?.....................    [197]
122.         .............................................    [197]
123.         .............................................    [197]

[                    210         220         230         240         250]
[                    .           .           .           .           .  ]

Homo01       ..CTG....C......C.GCC....A.....TAGA...T..........-    [248]
Homo02       ..CTG..........AC..CC....A.....TAG..............-    [248]
Ppan01       ..-...CC.C.........C...C.........................G    [248]
Ppan02       .C.................C.................T..........G    [250]
Ptve01       ..............C....C.T...........TA..............-    [246]
Ptve02       ..............C....CC.................T.....G....-    [246]
Ptve03       ..............C...??????????.........T.....G....-    [246]
Pttr01       ....G.............C.............................-    [246]
Pttr02       .G..G.............C.............................-    [246]
Pttr03       ....G...........G.............................T.-    [246]
1. (ref.)    AAACACAACTCCAAAGACATTCCTCCCCCACCCCGATACCAACAAACCT-    [246]
2.           .............................................-    [246]
3.           .............................................-    [246]
4.           ...................C.................C.?.....-    [246]
5.           ..?................C.............................-    [246]
```

```
6.          .................C.............................-   [246]
7.          .................C......T......................-   [246]
8.          .................C.............................-   [246]
9.          .................C.........................G....-   [246]
10.         .................C.............................-   [246]
11.         ................................T.........G....-   [246]
12.         ................................T..............-   [246]
13.         ................................T..............-   [246]
14.         ................................T..............-   [246]
15.         ............................................G....-   [246]
16.         ..............................................-   [246]
17.         .................C.............................-   [246]
18.         .................C.............................-   [246]
19.         ................................T..............-   [246]
20.         ................................T.........G....-   [246]
21.         ................................T..............-   [246]
22.         .................C............T..............-   [246]
23.         .................C............T..............-   [246]
24.         .................C............T..............-   [246]
25.         .................C............T..............-   [246]
26.         .................C............T..............-   [246]
27.         .................C............T..............-   [246]
28.         ................................AG..T.....G....-   [246]
29.         ................................AG..T..........-   [246]
30.         ................................AG..T..........-   [246]
31.         ..............?.................AG..T......?...-   [246]
32.         ............................T..A...T..........-   [246]
33.         ................................A...T..........-   [246]
34.         ................................A...T..........-   [246]
35.         ....................................T..........-   [246]
36.         ........T...........................T..........-   [246]
37.         ........T...........................T..........-   [246]
38.         ........T...........................T..........-   [246]
39.         ........T...........................T..........-   [246]
40.         ............................T...........G....-   [246]
41.         ....................................T..........-   [246]
42.         ....................................T..........-   [246]
43.         ....................................T..........-   [246]
44.         .................C..................T..........-   [246]
45.         ..............................................-   [246]
46.         .................C..................T..........-   [246]
47.         .................C..................T..........-   [246]
48.         ............................................G....-   [246]
49.         ............................................G....-   [246]
50.         ............................................G....-   [246]
51.         ............................................G....-   [246]
52.         ............G...............................G....-   [246]
53.         ............G...............................G....-   [246]
54.         ............................................G....-   [246]
55.         ............................................G....-   [246]
56.         ............................................G....-   [246]
57.         ..................C.........................G....-   [246]
58.         ..................C.............................-   [246]
59.         ............................................G....-   [246]
60.         ............................................G....-   [246]
61.         ............................................G....-   [246]
62.         ............................................G....-   [246]
```

```
63.  .....G....-  [246]
64.  ......T.....?G....-  [246]
65.  .....G....-  [246]
66.  .....G....-  [246]
67.  .....T.....G....-  [246]
68.  .....C.....-  [246]
69.  .....C.....-  [246]
70.  .....-  [246]
71.  .....T.....G....-  [246]
72.  .....T.....G....-  [246]
73.  .....G....-  [246]
74.  .....T.....G....-  [246]
75.  .....G....-  [246]
76.  .....G....-  [246]
77.  .....G....-  [246]
78.  .....G....-  [246]
79.  .....-  [246]
80.  .....-  [246]
81.  .....-  [246]
82.  .....-  [246]
83.  .....-  [246]
84.  .....-  [246]
85.  .....G....-  [246]
86.  .....-  [246]
87.  .....T.....-  [246]
88.  .....T.....-  [246]
89.  .....-  [246]
90.  .....-  [246]
91.  .....T.....G....-  [246]
92.  .....T.....G....-  [246]
93.  .....T.....G....-  [246]
94.  .....T.....G....-  [246]
95.  .....T.....G....-  [246]
96.  .....T.....G....-  [246]
97.  .....A.....G....-  [246]
98.  .....A.....G....-  [246]
99.  .....A.....G....-  [246]
100. .....G....-  [246]
101. .....G....-  [246]
102. .....G....-  [246]
103. .....G....-  [246]
104. .....T.....G....-  [246]
105. .....G....-  [246]
106. .....G....-  [246]
107. .....G....-  [246]
108. .....G....-  [246]
109. .....-  [246]
110. .....-  [246]
111. .....G....-  [246]
112. .....G....-  [246]
113. .....G....-  [246]
114. .....T.....T.....-  [246]
115. .....G....-  [246]
116. .....G....-  [246]
117. .....G....-  [246]
118. .....G....-  [246]
119. .....G....-  [246]
120. .....G....-  [246]
```

```
121.          ..?.........................?.............G....-   [246]
122.          ........T.....................................G....-   [246]
123.          ..............................T..........G....-   [246]

[                    260         270         280         290         300]
[                    .           .           .           .           .  ]

Homo01        .CC.GT................C.....A.GC.A.TT..CG.........   [298]
Homo02        .CC.A.................C.....A.GC.A.TT..CG.........   [298]
Ppan01        .C.G...-..............C.........TA..T..CG.........   [297]
Ppan02        .C.A...-..G.T.................G..A..T..CG.........   [299]
Ptve01        .CC.......G...AA.....A...........AC....CG.G.......   [296]
Ptve02        .CC.?.....G....A.................A.....CG.........   [296]
Ptve03        .CC.C...C.G....A.................A.....CG.........   [296]
Pttr01        .C...................C........C..C....C.G........   [296]
Pttr02        .C...................C........C..C....C.G........   [296]
Pttr03        .C.............C......C........C..C....C.G....G...   [296]
1. (ref.)     ATACTCCCTTAACAGTACATAGTACATACAATCGTACACTATACATAGCA
2.            ...................................................   [296]
3.            ...................................................   [296]
4.            ......T..............C........C......TC..........   [296]
5.            ......T..............C........C......TC..........   [296]
6.            ......T..............C.......GC......TC..........   [296]
7.            ......T..............C........C......TC..........   [296]
8.            ......T..............C........C......TC..........   [296]
9.            ......T..............C........C......TC..........   [296]
10.           ......T..............C........C......TC..........   [296]
11.           ......T.....................T...C.......C..........   [296]
12.           .....T.....G.....................C.......C..........   [296]
13.           .....TT.........................C.......C..........   [296]
14.           .....TT.........................C.......C..........   [296]
15.           ....C.T..............C........C.......C..........   [296]
16.           ....C.T..............C........C.......C..........   [296]
17.           ....C................C........C.......C..........   [296]
18.           ....C.T..............C........C.......C..........   [296]
19.           ......T..............C........C.......C..........   [296]
20.           ......T..............C........C......TC..........   [296]
21.           ......T..............C........C......TC..........   [296]
22.           ......T..............C........C...................   [296]
23.           ......T..............C........C.......C..........   [296]
24.           ......T..............C........C.......C..........   [296]
25.           ......T..............C........C.......C..........   [296]
26.           ......T..............C........C.......C..........   [296]
27.           ......T..............C........C.......C..........   [296]
28.           ..C..............................C.......C..........   [296]
29.           ..C..............................C.......C..........   [296]
30.           ..C..............................C.......C..........   [296]
31.           ..C..............................C.......C....?.....   [296]
32.           ..C...T..........................C.......C..........   [296]
33.           ..C...T..........................C.......C..........   [296]
34.           ..C...T..........................C.......C..........   [296]
35.           ..C...T..........................C.......C..........   [296]
36.           ..C...T..........................C.......C..........   [296]
37.           ..C...T..........................C.......C..........   [296]
38.           ..C...T........................GC.......C..........   [296]
39.           ..C...T..........................C.......C..........   [296]
```

```
40.      ......T.....................C.......C.....      [296]
41.      ..C...T.....................C.......C.....      [296]
42.      ..C...T.....................C.......C.....      [296]
43.      ..C...T.....................C.......C.....      [296]
44.      ..C...T.....................C.......C.....      [296]
45.      ......T.....................C.......C......G...  [296]
46.      ......T.....................C.......C.....      [296]
47.      ......T.....................C.......C.....      [296]
48.      ......T.....................C.......C.....      [296]
49.      ......T.....................C.......C.....      [296]
50.      ....................C.......C..C....C.....      [296]
51.      ....................C.......C..C....C.....      [296]
52.      G...........................C.......C.....      [296]
53.      G...........................C.......C.....      [296]
54.      G...........................C.......C.....      [296]
55.      G.....T.....................C.......C.....      [296]
56.      G.....T.....................C.......C.....      [296]
57.      G...........................C.......C.....      [296]
58.      G...........................C.......C.....      [296]
59.      G...........................C.......C.....      [296]
60.      G.....T.....................C.......C.....      [296]
61.      G...........................C.......C.....      [296]
62.      G...........................C.......C.....      [296]
63.      G...........................C......TC.....      [296]
64.      G...........................C.......C....?.....  [296]
65.      G.....................C.....C.......C.....      [296]
66.      ............................C.......C.....      [296]
67.      ......T.....................C.......C.....      [296]
68.      ............................C.............      [296]
69.      ............................C.......C.....      [296]
70.      .....T......................C.......C.....      [296]
71.      ....C.T.....................C......TC.....      [296]
72.      ....C.T.....................C.......C.....      [296]
73.      ......T.....................C.......C.....      [296]
74.      ......T.....................C.......C.....      [296]
75.      ......T.....................C.......C.....      [296]
76.      ......T...G.................C.......C.....      [296]
77.      ......T.....................C.......C.....      [296]
78.      ......T.....................C.......C.....      [296]
79.      ............................C.......C.....      [296]
80.      ......T.....................C.......C......G...  [296]
81.      ......T.....................C.......C.....      [296]
82.      ......T.....................C.......C.....      [296]
83.      ......T.....................C.......C.....      [296]
84.      ......T.....................C.......C.....      [296]
85.      ......T.....................C.......C.....      [296]
86.      .C....T.....................C..C....C.....      [296]
87.      ......T.....................C.......C.....      [296]
88.      ......T.....................C.......C.....      [296]
89.      ......T.....................C.......C.....      [296]
90.      ......T.....................C.......C.....      [296]
91.      ......T.....................TC......TC.....      [296]
92.      ......T.....................C.......C.....      [296]
93.      ......T.....................C.......C.....      [296]
94.      ......T.....................C.......C.....      [296]
95.      ......T.....................C......TC.....      [296]
96.      ......T.....................C.......C.....      [296]
```

```
97.          ......T.....................C.......C.........   [296]
98.          ......T.....................GC......C.........   [296]
99.          ......T.....................C.......C.........   [296]
100.         ......T.....................C.......C.........   [296]
101.         ......T.....................C.......C.........   [296]
102.         ......T.....................C.......C.........   [296]
103.         ......T.....................C.......C.........   [296]
104.         ......T.....................C.......C.........   [296]
105.         ......T.....................C.......C.........   [296]
106.         ......T............C........C.......C.........   [296]
107.         ......T.....................C.......C.........   [296]
108.         ......T.....................C.......C.........   [296]
109.         ......T.....................C.......C.........   [296]
110.         ......T.....................C.......C.........   [296]
111.         ......T.....................C.......C.........   [296]
112.         ......T.....................C.......C.........   [296]
113.         ......T.....................C.......C.........   [296]
114.         ..C...T.....................C.......C.........   [296]
115.         ......T.....................C.......C.........   [296]
116.         ......T.....................C.......C.........   [296]
117.         ............................C.......C.........   [296]
118.         ............................C.......C.........   [296]
119.         ............................C.......C.........   [296]
120.         ............................C.......C.........   [296]
121.         .C.........?........?.........?C.......C?.....   [296]
122.         G...........................C.......C.........   [296]
123.         ............................C.......C.........   [296]

[                  310         320         330         340         350]
[                  .           .           .           .           .  ]

Homo01       .......G........C....T.T.....T.....A.............   [348]
Homo02       .......G........C............T.....A.............   [348]
Ppan01       .......G..........C..................C..............   [347]
Ppan02       .......G.............................C........????..   [349]
Ptve01       .......G..........C..T.............AT...........G.   [346]
Ptve02       .......G......T.T.C....T...........C..............   [346]
Ptve03       .......G......T.T.C....T...........C..............   [346]
Pttr01       .....T.G..............T...........C..............   [346]
Pttr02       .......G..............T...........C..............   [346]
Pttr03       .......G..........C..T.............C..............   [346]
1. (ref.)    CATTACAATCAAATCCATTCTCGCCCCCACGGATGGCCCCCCTCAGATAG   [346]
2.           .......G.............................C.........A....   [346]
3.           ....................................C.........A....   [346]
4.           .......G.............................C??..????????G.   [346]
5.           .......G.............................C..............   [346]
6.           .......G.............................C..............   [346]
7.           .......G.............................C..............   [346]
8.           .......G.............................C..............   [346]
9.           .......G.............................C..............   [346]
10.          .......G.............................C..............   [346]
11.          .....G.G.............................C..........G...   [346]
12.          .......G.............................C..........G...   [346]
13.          .......G.............................C..............   [346]
14.          .......G.............................C.........AG...   [346]
15.          .......G..........C..................C..............   [346]
16.          .......G..........C..................C..............   [346]
```

```
17.     .......G.........C.................C.............   [346]
18.     .......G.........C.................C.............   [346]
19.     .......G.........C.................C.............   [346]
20.     .......G.........C.................C.............   [346]
21.     .......G.........C.................C.............   [346]
22.     .......G.........C.................C.............   [346]
23.     .......G.......T..C.................C.............   [346]
24.     .......G.......T..C.................C.............   [346]
25.     .......G.........C.................C.............   [346]
26.     ....G..G.........C.................C.............   [346]
27.     ....G..G.........C.................C.............   [346]
28.     .....T.G.........C.................C.............   [346]
29.     .....T.G.........C.................T.............   [346]
30.     .....T.G.........C.................C.............   [346]
31.     ....?T.G.........C.................C........?.....   [346]
32.     .....T.G.........C.................C.............   [346]
33.     .....T.G.........C.................C.............   [346]
34.     .....T.G.........C.................C.............   [346]
35.     .......G.........C.................C.............   [346]
36.     .......G.........C.................C.............   [346]
37.     .......G.........C....T............C.............   [346]
38.     .......G.........C.................C.............   [346]
39.     .......G.........C.................T.............   [346]
40.     .......G.........C.................C.............   [346]
41.     .......G.........C.................C.............   [346]
42.     .......G.........C.................C.............   [346]
43.     .......G.........C.................C.............   [346]
44.     .......G.........C.................C.............   [346]
45.     .......G.........C.................C.............   [346]
46.     ....T..G.........C.................C.............   [346]
47.     ....T..G...........................C.............   [346]
48.     .......G.........C.................C.............   [346]
49.     .......G.........C.................C.............   [346]
50.     .......G.........C.................T.............   [346]
51.     .......G.........C.................C.............   [346]
52.     .......G.........C.........G......C....T........   [346]
53.     .......G.........C.................C....T........   [346]
54.     .......G.........C.................C....T........   [346]
55.     .......G.........C.................C.............   [346]
56.     .......G.........C.................C.............   [346]
57.     .......G.........C.................C.............   [346]
58.     .......G.........C.................C.............   [346]
59.     .......G.........C.................C.............   [346]
60.     .......G.........C.................C.............   [346]
61.     .......G.........C.................C.............   [346]
62.     .......G.........C.................C.............   [346]
63.     .......G.........C.................C.............   [346]
64.     ....?..G.........C.................C........??....   [346]
65.     .......G.........C.................C.............   [346]
66.     .......G.........C.................C.............   [346]
67.     .......G.........C.................C.............   [346]
68.     .......G.........C.................C.............   [346]
69.     .......G.........C.................C...T.........   [346]
70.     .......G.........C.................C.............   [346]
71.     .......G.........C.................C.............   [346]
72.     .......G...........................C.............   [346]
73.     .......G.........C.................C.............   [346]
```

```
74.          .......G.........C...............C.............   [346]
75.          .......G.........C...............C.............   [346]
76.          .......G.........C...............C.............   [346]
77.          .......G.........C...............C.............   [346]
78.          .......G.........C...............C.............   [346]
79.          .......G.........C...............C.............   [346]
80.          .......G.........C...............C.............   [346]
81.          .......G.........C...............C.............   [346]
82.          .......G.........C...............C.............   [346]
83.          .......G.........C...............C.............   [346]
84.          .......G.........C...............C.............   [346]
85.          .......G.........C...............C.............   [346]
86.          .......G.........C...............C.............   [346]
87.          .......G.........C...............C.............   [346]
88.          .......G.........C...............C.............   [346]
89.          .......G.........C...............C.............   [346]
90.          .......G.........C...............C.............   [346]
91.          .......G.........C...............C.............   [346]
92.          .......G.........C...............C.............   [346]
93.          .......G.........C...............C.............   [346]
94.          .......G.........C...............C.............   [346]
95.          .......G.........C...............C.............   [346]
96.          .......G.........C...............C.............   [346]
97.          .......G.........C...............C.............   [346]
98.          .......G.........C...............C.............   [346]
99.          .......G.........C...............C.............   [346]
100.         .......G.........C...............C.............   [346]
101.         .......G.........C...............C.............   [346]
102.         .......G.........C...............C.............   [346]
103.         .......G.........C...............C.............   [346]
104.         .......G.........C...............C.............   [346]
105.         .......G.........C...............C.............   [346]
106.         .......G.........C...............C.............   [346]
107.         .......G.........C...............C.............   [346]
108.         .......G.........C...............C.............   [346]
109.         .......G.........C...............C.............   [346]
110.         .......G.........C...............C.............   [346]
111.         .......G.........C...............C.............   [346]
112.         .......G.........C...............C.............   [346]
113.         .......G.........C...............C.............   [346]
114.         .......G.........C...............C.............   [346]
115.         .......G.........C...............C.............   [346]
116.         .......G.........C...............C.............   [346]
117.         .......G.........C...............C.............   [346]
118.         .......G.........C...............T.............   [346]
119.         .......G.........C...............T.............   [346]
120.         .......G.........C...............C.............   [346]
121.         .......G.........C...............C.............   [346]
122.         .......G.........C...............C.............   [346]
123.         .......G.........C...............C.............   [346]

[                    360         370]
[                    .           .  ]

Homo01       ??????????????????????    [370]
Homo02       .AG....................    [370]
```

```
Ppan01        .A.....................   [369]
Ppan02        ..G.??????????????????   [371]
Ptve01        .A........A............   [368]
Ptve02        .A..??????????????????   [368]
Ptve03        .A..??????????????????   [368]
Pttr01        .AG.......TT..........   [368]
Pttr02        .AG.......CT..........   [368]
Pttr03        .AG.......CT..........   [368]
1. (ref.)     GGATCCCTTGGCCACCATCCTC   [368]
2.            ..........C...........   [368]
3.            ..........C...........   [368]
4.            ..G.??????????????????   [368]
5.            ..G.??????????????????   [368]
6.            ..G.......CT..........   [368]
7.            ..G.......CT..........   [368]
8.            ..G.......CT..........   [368]
9.            ..G.......CT..........   [368]
10.           ..G.......CT..........   [368]
11.           ..G.......C...........   [368]
12.           ..G.......C...........   [368]
13.           ..G.......CT..........   [368]
14.           ..G.......C...........   [368]
15.           ..G.......C...........   [368]
16.           ..G.......C...........   [368]
17.           ..G.......C...........   [368]
18.           ..G.??????????????????   [368]
19.           ..G.......C...........   [368]
20.           ..G.......C...........   [368]
21.           ..G.......C...........   [368]
22.           ..G.......C...........   [368]
23.           ..G.......C...........   [368]
24.           ..G.......C...........   [368]
25.           ..G.??????????????????   [368]
26.           ..G.......C...........   [368]
27.           ..G.......C...........   [368]
28.           ..G.......C...........   [368]
29.           ..G.......C...........   [368]
30.           ..G.......C...........   [368]
31.           ..G.??????????????????   [368]
32.           ..G.??????????????????   [368]
33.           ..G.......C...........   [368]
34.           ..G.......C...........   [368]
35.           ..G.......C...........   [368]
36.           ..G.......C...........   [368]
37.           ..G.......C...........   [368]
38.           ..G.......C...........   [368]
39.           ..G.......C...........   [368]
40.           ..G.......C...........   [368]
41.           ..G.??????????????????   [368]
42.           ..G.??????????????????   [368]
43.           ..G.......C...........   [368]
44.           ..G.......C...........   [368]
45.           ..G.......C...........   [368]
46.           ..G.......C...........   [368]
47.           ..G.......C...........   [368]
48.           ..G.......CT..........   [368]
49.           ..G.......C...........   [368]
50.           ..G.......C...........   [368]
```

```
51.         ..G.......C.....    [368]
52.         ..G.......CT....    [368]
53.         ..G.......CT....    [368]
54.         ..G.......CT....    [368]
55.         ..G.......CT....    [368]
56.         ..G.......CT....    [368]
57.         ..G.......C.....    [368]
58.         ..G.......C.....    [368]
59.         ..G.......C.....    [368]
60.         ..G.......C.....    [368]
61.         ..G.......C.....    [368]
62.         ..G.......C.....    [368]
63.         ..G.......C.....    [368]
64.         ..G.??????????????????    [368]
65.         ..G.......C.....    [368]
66.         ..G.......C.....    [368]
67.         ..G.......C.....    [368]
68.         ..G.......C.....    [368]
69.         ..G.......C.....    [368]
70.         ..G.......C.....    [368]
71.         ..G.......C.....    [368]
72.         ..G.......C.....    [368]
73.         ..G.......C.....    [368]
74.         ..G.......C.....    [368]
75.         ..G.......C.....    [368]
76.         ..G.......C.....    [368]
77.         ..G.......C.....    [368]
78.         ..G.......C.....    [368]
79.         ..G.......C.....    [368]
80.         ..G.......C.....    [368]
81.         ..G.......C.....    [368]
82.         ..G.......C.....    [368]
83.         ..........C.....    [368]
84.         ..G.......C.....    [368]
85.         ..G.......C.....    [368]
86.         ..G.......C.....    [368]
87.         ..G.......T.....    [368]
88.         ..G.......C.....    [368]
89.         ..G.......C.....    [368]
90.         ..G.......C.....    [368]
91.         ..G.......C.....    [368]
92.         .AG.......C.....    [368]
93.         ..G.......C.....    [368]
94.         ..G.......C.....    [368]
95.         ..G.......C.....    [368]
96.         ..G.......C.....    [368]
97.         ..G.......C.....    [368]
98.         ..G.......C.....    [368]
99.         ..G.......C.....    [368]
100.        ..G.......C.....    [368]
101.        ..G.......C.....    [368]
102.        ..G.......C.....    [368]
103.        ..G.......C.....    [368]
104.        ..G.......C.....    [368]
105.        ..G.......T.....    [368]
106.        ..G.......C.....    [368]
107.        ..G.......C.....    [368]
```

```
108.        ..G.......C..........   [368]
109.        ..G.......C..........   [368]
110.        ..G.......C..........   [368]
111.        ..G.......C..........   [368]
112.        ..G.......C..........   [368]
113.        ..G.......C..........   [368]
114.        ..G.......C..........   [368]
115.        ..G.......C..........   [368]
116.        ..G.......C..........   [368]
117.        ..G.......C..........   [368]
118.        ..G.......C..........   [368]
119.        ..G.......C..........   [368]
120.        ..G.......C..........   [368]
121.        ..G.??????????????????   [368]
122.        ..G.......C..........   [368]
123.        ..G.......C..........   [368]
```

Note:

1. Outgroups are described in Appendix 1. Ingroup sequences (123 *Pan troglodytes schweinfurthii* sequences) are described by number in Appendices 1 and 3. Sequences taken from Morin *et. al.* (1994) are were aligned by hand to the sequences generated in the present study. Missing data in the Morin sequences are coded above as "?". The first *P. t. schweinfurthii* sequence (taxon label: "1. (ref.)") was used as the reference sequence. Dots indicate sequence identity with the reference sequence.

Appendix 3: List of 123 eastern chimpanzee haplotypes

Haplotype [1]	Found in populations [2]	Number of individuals with haplotype [3]
1	IIAA	3
2	IIAA	2
3	EEZO, IIAA	1,1
4	GEKA	3
5	GEKA	3
6	RIKA	2
7	RIKA	1
8	KEDN	5
9	IARA	1
10	EEZO, GEKA, RIKA	2,1,2
11	IIAA	1
12	KENO	3
13	IIAA	1
14	BOPI	1
15	IILA	3
16	IILA	3
17	KEKU, SINI	1,3
18	GEKA	2
19	NESN	3
20	IILA	2
21	IILA	1
22	KEDN	1
23	IILA	1
24	IIAA	1
25	GEKA	2
26	EEZO	1
27	BAMA	1
28	SIMU	1
29	SIMU	1
30	RIKA, SIMU	1,3
31	GEKA	2
32	GEKA	2
33	KUSL	1
34	EEZO, RIKA	1,1
35	NESN	2
36	KEKU, SINI	2,1
37	EEZO, KENO	1,1
38	BOSO	1
39	BOSO	1
40	BOSO, IARA	1,1
41	GEKA	1

Haplotype	Found in populations	Number of individuals with haplotype
42	GEKA	1
43	NESN	2
44	BAMA, BOPI, BOSO	1,1,4
45	IIAW	2
46	NESN	2
47	NESN	2
48	IIAE	7
49	KUSL	1
50	SIMU	1
51	EEZO, IIAW, KEKU, SIMU, SINI	1,3,1,7,3
52	KUSL	1
53	KUSL	1
54	KUSL	1
55	KUSL	1
56	KUSL	1
57	IIAW	2
58	IIAW	2
59	IILA	1
60	IILA	1
61	IIAA	1
62	IIAE	2
63	IIAA	1
64	GEKA	1
65	EEZO	1
66	IIAW	1
67	TOBA	3
68	TOBA	2
69	TOBA	1
70	EEZO	1
71	IIAE	1
72	IIAA	1
73	RIKA,TOBA	1,1
74	IARA	1
75	EEZO, IARA, KEKA	2,3,2
76	EEZO	1
77	EEZO	1
78	EEZO	1
79	NESN	1
80	IILA	1
81	EEZO	1
82	BOPI, BOSO, EEZO	1,4,2
83	BOPI	1
84	BOPI, KEKU, RIKA	3,1,1
85	TOBA	1

Haplotype	Found in populations	Number of individuals with haplotype
86	EEZO	1
87	IIAW	1
88	IIAW	1
89	EEZO,IIAW,KEDN,KEKU,KENO,RIKA	1,1,1,1,2,1
90	BOPI	1
91	IARA	1
92	SINI	2
93	RIKA	1
94	KEKA	1
95	IARA	1
96	EEZO, KEDN, KEKA, KUSL	1,6,2,2
97	KEKA, KENO, RIKA	4,2,1
98	BOSO	1
99	BOSO	1
100	SINI	1
101	EEZO, IILA, KUSL, RIKA	2,1,1
102	BAMA, IARA, KEKU, SINI	1,3,2,1
103	EEZO	1
104	TOBA	1
105	EEZO	1
106	KEKA, KEKU	3,1
107	EEZO	1
108	KUSL	1
109	BOPI	1
110	IILA	1
111	SINI	1
112	NESN	1
113	KEDN, KEKA, KUSL	6,2,2
114	BOPI, BOSO, IARA	3,1,1
115	BAMA, BOSO, EEZO, KUSL	8,1,2,2
116	BAMA, KEKA	1,1
117	TOBA	1
118	IIAE	1
119	IILA	1
120	IIAE	1
121	GEKA	1
122	IIAA	1
123	BAMA	1

Notes

1. Haplotypes are arbitrarily numbered. Numbers cross-reference Appendices 1 and 2.
2. See Table 2.1 and associated text for description of populations
3. Numbers separated by commas refer to the number of individuals posessing the haplotype in each respective sampling location listed.

References

Albrecht, H. and S. C. Dunnett, 1971. *Chimpanzees in Western Africa.* Munich, Piper Verlag.

Allendorf, F. W., 1983. Isolation, gene flow, and genetic differentiation among populations. In *Genetics and Conservation: A Reference for Managing Wild Animal and Plant Populations*, edited by Schonewald-Cox, C. M., S. M. Chambers, B. MacBryde and W. L. Thomas, p. 51-65. Menlo Park, The Benjamin/Cummings Publishing Company, Inc.

Allendorf, F. W. and R. F. Leary, 1986. Heterozygosity and fitness in natural populations of animals. In *Conservation Biology: The Science of Scarcity and Diversity*, edited by Soulé, M. E., p. 57-76. Sunderland, Sinauer Associates, Inc.

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. de Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden and I. G. Young, 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290: 457-465.

Applied Biosystems, Inc., 1993. Taq DyeDeoxy Terminator Cycle Sequencing Kit, part #901497. Foster City, CA.

Applied Biosystems, Inc., 1994a. 373 DNA Sequencing System: User's Manual. Foster City, CA.

Applied Biosystems, Inc., 1994b. *AutoAssembler™: DNA sequence assembly software.* Version 2.1. Foster City, CA.

Arnason, E., S. Pálsson and A. Arason, 1992. Gene flow and lack of population differentiation in Atlantic cod, *Gadus morhua* L., from Iceland, and comparison of cod from Norway and Newfoundland. *J. Fish Biol.* 40: 751-770.

Atchley, W. R. and W. M. Fitch, 1991. Gene trees and the origins of inbred strains of mice. *Science* 254: 554-558.

Atkinson, I., 1989. Introduced animals and extinctions. In *Conservation for the Twenty-first Century*, edited by Western, D. and M. C. Pearl, p. 54-69. New York, Oxford University Press.

Aubréville, A., 1949. *Climats, forêts et désertification de l'Afrique tropicale.* Paris, Société d'Éditions Géographiques, Maritimes et Coloniales.

Avise, J. and W. S. Nelson, 1989. Molecular genetic relationships of the extinct Dusky Seaside Sparrow. *Science* 243: 646-648.

Avise, J. C., 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution* 43: 1192-1208.

Avise, J. C., 1992. Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos* 63: 62-76.

Avise, J. C., 1994. *Molecular Markers, Natural History and Evolution.* New York, Chapman and Hall.

Avise, J. C., J. Arnold, R. M. Ball, e. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb and N. C. Saunders, 1987. Intraspecific Phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.* 18: 489-522.

Ayres, J. M. and T. H. Clutton-Brock, 1992. River boundaries and species range size in Amazonian primates. *Am. Nat.* 140: 531-537.

Baker, A. J., M. D. Denison, A. Lynch and G. L. Grand, 1990a. Genetic divergence in peripherally isolated populations of chaffinches in the Atlantic islands. *Evolution* 44: 981-999.

Baker, C. S., S. R. Palumbi, R. H. Lambertsen, M. T. Weinrich, J. Calambokidis and S. J. O'Brien, 1990b. Influence of seasonal migration on geographic distribution of mitochondrial DNA haplotypes in humpback whales. *Nature* 344: 238-240.

Baldwin, P. J., W. C. McGrew and C. E. G. Tutin, 1982. Wide-ranging chimpanzees at Mt. Assirik, Senegal. *IntJ. Primatol.* 3: 367-385.

Baldwin, P. J., J. S. Pí, W. C. McGrew and C. E. G. Tutin, 1981. Comparison of nests made by different populations of chimpanzees (*Pan troglodytes*). *Primates* 22: 474-486.

Ball, R. M., S. Freeman, F. C. James, E. Bermingham and J. C. Avise, 1988. Phylogeographic population structure of red-winged blackbirds assessed by mitochondrial DNA. *Proc. Nat. Acad. Sci., USA.* 85: 1558-1562.

Barbujani, G., 1987. Autocorrelation of gene frequencies under isolation by distance. *Genetics* 117: 777-782.

Barbujani, G., G. Bertorelle, G. Capitani and R. Scozzari, 1995. Geographical structuring in the mtDNA of Italians. *Proc. Nat. Acad. Sci., USA* 92: 9171-9175.

Barbujani, G., A. Pilastro, S. De Domenico and C. Renfrew, 1994. Genetic variation in North Africa and Eurasia: Neolithic demic diffusion vs. Paleolithic colonization. *Am. J. Phys. Anthro.* 95: 137-154.

Beardmore, J. A., 1983. Extinction, survival and genetic variation. In *Genetics and Conservation: A Reference for Managing Wild Animal and Plant Populations*, edited by Schonewald-Cox, C. M., S. M. Chambers, B. MacBryde and W. L. Thomas, p. 125-151. Menlo Park, The Benjamin/Cummings Publishing Company, Inc.

Bermingham, E. and J. C. Avise, 1986. Molecular zoogeography of freshwater fishes in the south-eastern United States. *Genetics* 113: 939-965.

Bertorelle, G. and G. Barbujani, 1995. Analysis of DNA diversity by spatial autocorrelation. *Genetics* 140: 811-819.

Bonnefille, R. and G. Riollet, 1988. The Kashiru pollen sequence (Burundi) palaeoclimatic implications for the last 40,000 yr B.P. in tropical Africa. *Quat. Res.* 30: 19-35.

Bonnefille, R., J. C. Roeland and J. Guiot, 1990. Temperature and rainfall estimates for the past 40,000 years in equatorial Africa. *Nature* 346: 347-349.

Bonnell, M. L. and R. K. Selander, 1974. Elephant seals: genetic variation and near extinction. *Science* 184: 908-909.

Booth, A. H., 1954. The Dahomey gap and the mammalian fauna of the west African forests. *Rev. Zool. Bot. Afr.* 50: 3-4.

Booth, A. H., 1957. The Niger, the Volta and the Dahomey gap as geographic barriers. *Evolution* 12: 48-62.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd and L. L. Cavalli-Sforza, 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368: 455-458.

Bray, J. R., 1977. Pleistocene volcanism and glacial initiation. *Science* 197: 251-254.

Brown, J. H., 1988. Species diversity. In *Analytical Biogeography*, edited by Myers, A. A. and P. S. Giller, p. 57-89. London, Chapman & Hall.

Brown, W. M., 1980. Polymorphism in mitochondrial DNA of humans revealed by restriction endonuclease analysis. *Proc. Nat. Acad. Sci., USA* 77: 3605-3609.

Brundin, L. Z., 1988. Phylogenetic biogeography. In *Analytical Biogeography*, edited by Myers, A. A. and P. S. Giller, p. 343-369. London, Chapman & Hall.

Burke, K., A. B. Durotoye and A. J. Whiteman, 1971. A dry phase south of the Sahara 20,000 years ago. *W. Afr. J. Arch.* 1: 1-8.

Burton, R. S. and B.-N. Lee, 1994. Nuclear and mitochondrial gene genealogies and allozyme polymorphism across a major phylogeographic break in the copepod *Tigriopus californicus*. *Proc. Nat. Acad. Sci., USA* 91: 5197-5201.

Butynski, T. M., 1985. Primates and their conservation in the Impenetrable (Bwindi) Forest, Uganda. *Primate Cons.* 6: 68-72.

Butzer, K. W., G. L. Isaac, J. L. Richardson and C. Washbourn-Kamau, 1972. Radiocarbon dating of East African lake levels. *Science* 175: 1069-1076.

Caccone, A. and J. R. Powell, 1989. DNA divergence among hominoids. *Evolution* 43: 925-942.

Cairns, S. J. and S. J. Schwager, 1987. A comparison of association indices. *Anim. Behav.* 35: 1454-1469.

Cann, R. L., M. Stoneking and A. C. Wilson, 1987. Mitochondrial DNA and human evolution. *Nature* 325: 31-36.

Carcasson, R. H., 1964. A preliminary survey of the zoogeography of African butterflies. *E. Afr. Wildlife J.* 2: 122-157.

Carr, S. M., S. W. Ballinger, J. N. Derr, L. H. Blankenship and J. W. Bickham, 1986. Mitochondrial DNA analysis of hybridization between sympatric white-tailed deer and mule deer in west Texas. *Proc. Nat. Acad. Sci., USA.* 83: 9576-9580.

Carson, H. L., 1983. The genetics of the founder effect. In *Genetics and Conservation: A Reference for Managing Wild Animal and Plant Populations*, edited by Schonewald-Cox, C. M., S. M. Chambers, B. MacBryde and W. L. Thomas, p. 189-200. Menlo Park, The Benjamin/Cummings Publishing Company, Inc.

Cavalli-Sforza, L. L. and A. W. F. Edwards, 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* 21: 550-570.

Cerling, T. E., R. L. Hay and J. R. O'Neil, 1977. Isotopic evidence for dramatic climatic changes in East Africa during the Pleistocene. *Nature* 267: 137-138.

Chakraborty, R. and K. K. Kidd, 1991. The utility of DNA typing in forensic work. *Science* 254: 1735-1739.

Chakraborty, R. and M. Nei, 1977. Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution* 31: 347-356.

Chapman, C. A., 1983. Speciation of tropical rainforest primates of Africa: insular biogeography. *Afr. J. Ecol.* 21: 297-308.

Chapman, C. A. and F. White, 1995. Nearest neighbour distances in chimpanzees and bonobos. *Folia Primatol.* 63: 181-191.

Chapman, C. A. and R. W. Wrangham, 1993. Range use of the forest chimpanzees of Kibale: implications for the understanding of chimpanzee social organization. *Am. J. Primatol.* 31: 263-273.

Cliff, A. D. and J. K. Ord, 1981. *Spatial Processes.* London, Pion.

CLIMAP Project Members, 1976. The surface of the ice-age Earth. *Science* 191: 1131-1137.

Cockerham, C. C., 1969. Variance of gene frequencies. *Evolution* 23: 72-84.

Coetzee, J. A., 1964. Evidence for a considerable depression of the vegetation belts during the upper Pleistocene on the East African mountains. *Nature* 204: 564-566.

Coetzee, J. A., 1987. Palynological intimations on the East African mountains. *Palaeoecology of Africa* 18: 231-244.

COHMAP Members, 1988. Climatic changes of the last 18,000 years: observations and model simulations. *Science* 241: 1043-1052.

Colyn, M., 1991. L'Importance zoogéographique du bassin du fleuve Zaïre pour la spéciation: le cas des primates simiens. *Annales de la Musee Royale de l'Afrique Centrale, Tervuren, Belgique* 264: I-249.

Colyn, M. M., 1987. Les primates des forêts ombrophiles de la cuvette du Zaïre: interprétations zoogéographiques des modèles de distribution. *Rev. Zool. Afr.* 101: 183-

196.

Colyn, M. M., A. Gautier-Hion and W. Verheyen, 1991. A re-appraisal of paleoenvironmental history in Central Africa: evidence for a major fluvial refuge in the Zaire Basin. *J. Biogeog.* 18: 403-407.

Coolidge, H. J., 1929. A revision of the genus *Gorilla*. *Mem. Mus. Comp. Zool. Harv.* 50: 293-381.

Croizat, L., 1958. *Panbiogeography*. Caracas, published by the author.

Croizat, L., 1964. *Space, Time, Form: The Biological Synthesis*. Caracas, published by the author.

Croizat, L., 1982. Vicariance/vicariism, panbiogeography, "vicariance biogeography," etc.: a clarification. *Sys. Zool.* 31: 291-304.

Cronin, M. A., S. C. Amstrup, G. W. Garner and E. R. Vyse, 1991. Interspecific and intraspecific mitochondrial DNA variation in North American bears (*Ursus*). *Can. J. Zool.* 69: 2985-2992.

Cronin, M. A., M. E. Nelson and D. F. Pac, 1991. Spatial heterogeneity of mitochondrial DNA and allozymes among populations of white-tailed deer and mule deer. *J. Heredity* 82: 118-127.

Crowley, T. J., K.-Y. Kim, J. G. Mengel and D. A. Short, 1992. Modeling 100,000-year climate fluctuations in pre-Pleistocene time series. *Science* 255: 705-707.

Curie-Cohen, M., D. Yoshihara, L. Luttrell, K. Benforado, J. W. MacCluer and W. H. Stone, 1983. The effects of dominance on mating behavior and paternity in a captive group of rhesus monkeys (*Macaca mulatta*). *Am. J. Primatol.* 5: 127-138.

de Ruiter, J. R., W. Scheffrahn, G. J. J. M. Trommelen, A. G. Uitterlinden, R. D. Martin and J. A. R. A. M. v. Hoof, 1992. Male social rank and reproductive success in wild long-tailed macaques: paternity exclusion by blood protein analysis and DNA fingerprinting. In *Paternity in Primates: Genetic Tests and Theories*, edited by Martin, R. D., A. F. Dixson and E. J. Wickings, p. 175-191. Basel, Karger.

de Waal, F. B. M., 1982. *Chimpanzee Politics*. New York, Harper and Row.

de Waal, F. B. M., 1994. Chimpanzee's adaptive potential. In *Chimpanzee Cultures*, edited by Wrangham, R. W., W. C. McGrew, F. B. M. de Waal and P. G. Heltne, p. 243-260. Cambridge, Harvard University Press.

Degnan, S. M. and C. Moritz, 1992. Phylogeography of mitochondrial DNA in two species of white-eyes in Australia. *The Auk* 109: 800-811.

deMenocal, P. B., 1995. Plio-Pleistocene African climate. *Science* 270: 53-60.

Denevan, W. M. and C. Padoch, 1988. Swidden-fallow agroforestry in the Peruvian Amazon. *Adv. Econ. Botany* 5: 1-107.

Densmore, L. D., C. C. Moritz, J. W. Wright and W. M. Brown, 1989. Mitochondrial DNA analyses and the origin and relative age of parthenogenic lizards (genus *Cnemidophorus*). IV. Nine *sexlineatus*-group unisexuals. *Evolution* 43: 969-983.

Deuse, P., 1966. *Contribution a l'étude des tourbieres du Rwanda et du Burindi.* Report to Inst. Nat. Recherche Scient. Butare, Rwanda.

Di Rienzo, A. and A. C. Wilson, 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Nat. Acad. Sci., USA* 88: 1597-1601.

Diamond, A. W., 1979. The distribution of forest passerine birds and Quarternary climatic change in tropical Africa. *J. Zool. Soc., Lond.* 191: 379-402.

Diamond, J., 1989. Overview of recent extinctions. In *Conservation for the Twenty-first Century*, edited by Western, D. and M. C. Pearl, p. 37-41. New York, Oxford University Press.

Disotell, T. R., 1994. Generic level relationships of the Papionini (Cercopithecoidea). *Am. J. Phys. Anthro.* 94: 47-57.

Dixson, A. F., G. Anzenberger, M. A. O. Monterio Da Cruz, I. Patel and A. J. Jeffreys, 1992. DNA fingerprinting of free-ranging groups of common marmosets (*Callithrix jacchus jacchus*) in NE Brazil. In *Paternity in Primates: Genetic Tests and Theories*, edited by Martin, R. D., A. F. Dixson and E. J. Wickings, p. 192-202. Basel, Karger.

Dobson, A. P. and R. M. May, 1986. Disease and conservation. In *Conservation Biology: The Science of Scarcity and Diversity*, edited by Soulé, M. E., p. 345-365. Sunderland, Sinauer Associates, Inc.

Douglas, A. M., A. M. Georgalis and B. A. Atchison, 1993. Direct sequencing of double stranded PCR products incorporating a chemiluminescent detection procedure. *BioTechniques* 13: 132-134.

Downie, C. and P. Wilkinson, 1972. *The Geology of Kilimanjaro.* Sheffield, University of Sheffield Press.

Dufour, D. L., 1990. Use of tropical rainforests by native Amizonians. *Bioscience* 40: 652-659.

Duvall, S. W., I. S. Bernstein and T. P. Gordon, 1976. Paternity and status in a rhesus monkey group. *J. Repro. Fertil.* 47: 25-31.

Edwards, S. V., 1993. Mitochondrial gene genealogy and gene flow among island and mainland populations of a sedentary songbird, the grey-crowned babbler (*Pomatostomus temporalis*). *Evolution* 47: 1118-1137.

Edwards, S. V. and A. C. Wilson, 1990. Phylogenetically informative length polymorphism and sequence variability in the mitochondrial DNA of Australian songbirds (*Pomatostomus*). *Genetics* 126: 695-711.

Eglinton, G., S. A. Bradshaw, A. Rosell, M. Sarnthein, U. Pflaumann and R. Tiedemann, 1992. Molecular record of secular sea surface temperature changes on 100-year timescales for glacial terminations I, II and IV. *Nature* 356: 423-426.

Ehrlich, P. R., 1983. Genetics and the extinction of butterfly populations. In *Genetics and Conservation: A Reference for Managing Wild Animal and Plant Populations*, edited by Schonewald-Cox, C. M., S. M. Chambers, B. MacBryde and W. L. Thomas, p. 152-163. Menlo Park, The Benjamin/Cummings Publishing Company, Inc.

Elenga, H., A. Vincens and D. Schwartz, 1991. Presende d'elements forestiers montagnards sur les Plateaux Batéké (Congo) au Pléistocene supérieur: nouvelles données palynologiques. *Palaeoecology of Africa* 22: 239-244.

Ely, J., R. Deka, R. Chakraborty and R. E. Ferrell, 1992. Comparison of five VNTR loci between humans and chimpanzees. *Genomics* 14: 692-698.

Ely, J. and R. E. Ferrell, 1990. DNA "fingerprints" and paternity ascertainment in chimpanzees (*Pan troglodytes*). *Zoo Biology* 9: 91-98.

Emliani, C., 1955. Pleistocene temperatures. *J. Geol.* 63: 538-578.

Emliani, C., 1966. Paleotemperature analysis of Caribbean cores P 6304-8 and P 6304-9 and a generalised temperature curve for the last 425,000 years. *J. Geol.* 74: 109-126.

Endler, J. A., 1982. Pleistocene forest refuges: fact or fancy? In *Biological Diversification in the Tropics*, edited by Prance, G. T., p. 641-657. New York, Columbia University Press.

Erlich, H. A., Ed., 1989. *PCR Technology: Principles and Applications for DNA Amplification.* New York, Stockton Press.

Excoffier, L., P. E. Smouse and J. M. Quattro, 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

Farris, J. S., 1983. The logical basis of phylogenetic analysis. In *Advances in Cladistics*, edited by Platnick, N. and V. Funk, p. 7-36. New York, Columbia University Press.

Farris, J. S., 1986. *HENNIG86*. State University of New York at Stony Brook, Department of Ecology and Evolution.

Farris, J. S., 1989. The retention index and the rescaled consistency index. *Cladistics* 5: 417-419.

Fedigan, L. M., 1983. Dominance and reproductive success in primates. *Yearbook Phys. Anthro.* 26: 91-129.

Felsenstein, J., 1978. The number of evolutionary trees. *Sys. Zool.* 27: 27-33.

Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783-791.

Felsenstein, J., 1988. Phylogenies from molecular sequences: inferences and reliability. *Ann. Rev. Genet.* 22: 521-565.

Felsenstein, J., 1990. *PHYLIP: Phylogenetic Inference Package.* Version. 3.3. Department of Genetics, University of Washington, Seattle.

Ferris, S. D., A. C. Wilson and W. M. Brown, 1981. Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. *Proc. Nat. Acad. Sci., USA* 78: 2432-2436.

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

Flint, J., A. J. Boyce, J. J. Martinson and J. B. Clegg, 1989. Population bottlenecks in Polynesia revealed by minisatellites. *Hum. Genet.* 83: 257-263.

Foley, R., 1987. *Another Unique Species.* New York, John Wiley and Sons, Inc.

Fredrickson, W. T. and G. Sackett, 1984. Kin preferences in primates (*Macacca nemestrina*): relatedness or familiarity? *J. Comp. Psych.* 98: 29-34.

Fruth, B. and G. Hohmann, 1994a. Comparative analyses of nest building behavior in bonobos and chimpanzees. In *Chimpanzee Cultures*, edited by Wrangham, R. W., W. C. McGrew, F. B. M. de Waal and P. G. Heltne, p. 109-128. Cambridge, Harvard University Press.

Fruth, B. and G. Hohmann, 1994b. Nests: living artifacts of recent apes? *Current Anthro.* 35: 310-311.

Fucharoen, S., G. Fucharoen, P. Fucharoen and Y. Fukumaki, 1989. A novel ochre mutation in the ß-thallisemia gene of a Thai. *J. Biol. Chem.* 264: 7780-7783.

Fukuda, M., S. Wakasugi, T. Tsuzuki, H. Nomiyama and K. Shimada, 1985. Mitochondrial DNA-like sequences in the human nuclear genome. *J. Mol. Bio.* 186: 257-266.

Gallup, C. D., R. L. Edwards and R. G. Johnson, 1994. The timing of high sea levels over the past 200,000 years. *Science* 263: 796-800.

Garner, K. J. and O. A. Ryder, 1992. Some applications of PCR to studies in wildlife genetics. *Symp. Zool Soc. Lond.* 64: 167-181.

Gasse, F., R. Téhet, A. Durand, E. Gibert and J.-C. Fontes, 1990. The arid-humid transition in the Sahara and the Sahel during the last deglaciation. *Nature* 346: 141-146.

Geary, R. C., 1954. The contiguity ratio and statistical mapping. *Incorp. Stat.* 5: 115-145.

Ghiglieri, M. P., 1984. *The chimpanzees of Kibale Forest: a field study of ecology and social structure.* New York, Columbia University Press.

Ghiglieri, M. P., 1989. Hominoid sociobiology and hominid social evolution. In *Understanding Chimpanzees*, edited by Heltne, P. G. and L. A. Marquardt, p. 370-379. Cambridge, Harvard University Press.

Gilbert, D. A., N. Lehman, S. J. O'Brien and R. K. Wayne, 1990. Genetic fingerprinting reflects population differentiation in the California Channel Island fox. *Nature* 344: 764-766.

Gilbert, D. A., C. Packer, A. E. Pusey, J. C. Stephens and S. J. O'Brien, 1991. Analytical DNA fingerprinting in lions: parentage, genetic diversity and kinship. *J. Heredity* 82: 378-386.

Gilpin, M. E. and M. E. Soulé, 1986. Minimum viable populations: processes of species extinction. In *Conservation Biology: The Science of Scarcity and Diversity*, edited by Soulé, M. E., p. 19-34. Sunderland, Sinauer Associates, Inc.

Ginsberg, J. R. and T. P. Young, 1992. Measuring association between individuals or groups in behavioural studies. *Anim. Behav.* 44: 377-379.

Goodall, J., 1962. Nest building behavior in the free-ranging chimpanzee. *Ann. New York Acad. Sci.* 102: 455-467.

Goodall, J., 1986. *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge, Harvard University Press.

Goodall, J., A. Bandora, E. Bergmann, C. Busse, H. Matama, E. Mpongo, A. Pierce and D. Riss, 1979. Inter-community interactions in the chimpanzee population of the Gombe National Park. In *The Great Apes*, edited by Hamburg, D. A. and E. R. McCown, p. 13-53. Menlo Park, Benjamin Cummings.

Goodenow, M., T. Huet, W. Saurin, S. Kwok, J. Sninsky and S. Wain-Hobson, 1989. HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *JAIDS* 2: 344-352.

Gray, I. C. and A. J. Jeffreys, 1989. Evolutionary transience of hypervariable minisatellites in man and the primates. 241-253.

Grove, A. T. and A. Warren, 1968. Quaternary landforms and climate on the south side of the Sahara. *Geog. J*. 134: 194-208.

Groves, C. P., 1970. Population systematics of the gorialla. *J. Zool. Lond.* 161: 287-300.

Groves, C. P., 1971. Distribution and place of origin of the gorilla. *Man* 6: 44-51.

Grubb, P., 1982. Refuges and dispersal in the speciation of African forest mammals. In *Biological Diversification in the Tropics*, edited by Prance, G. T., p. 537-553. New York, Columbia University Press.

Guiot, J., A. Pons, J. L. de Beaulieu and M. Reille, 1989. A 140,000-year continental climate reconstruction from two European pollen records. *Nature* 338: 309-313.

Haddow, A. J., 1952. Field and laboratory studies on an African monkey, *Cercopithecus ascanius schmidti* Matschie. *Proc. Zool. Soc. Lond.* 122: 297-394.

Haffer, J., 1982. General aspects of the refuge theory. In *Biological Diversification in the Tropics*, edited by Prance, G. T., p. 6-24. New York, Columbia University Press.

Hale, L. R. and R. S. Singh, 1991. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. IV. Mitochondrial DNA variation and the role of history vs. selection in the genetic structure of geographic populations. *Genetics* 129: 102-117.

Hamilton, A., 1976. The significance of patterns of distribution shown by forest plants and animals in tropical Africa for the reconstruction of upper Pleistocene paleoenvironments: a review. *Palaeoecology of Africa* 9: 63-97.

Hamilton, A., 1987. Vegetation and climate of Mt. Elgon during the late Pleistocene and Holocene. *Palaeoecology of Africa* 18: 283-304.

Hamilton, A., D. Taylor and J. C. Vogel, 1986. Early forest clearance and environmental degredation in south-west Uganda. *Nature* 320: 164-167.

Hamilton, A. C., 1981. The quarternary history of African forests: its relevance to conservation. *African J. Ecol.* 19: 1-6.

Hamilton, A. C., 1982. *Environmental History of East Africa: A Study of the Quaternary*. London, Academic Press.

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

Hamilton, A. C., 1988. Guenon evolution and forest history. In *A Primate Radiation: Evolutionary Biology of the African Guenons*, edited by Gautier-Hion, A., F. Bourliere, J.-P. Gautier and J. Kingdon, p. 13-34. Cambridge, Cambridge University Press.

Hamilton, A. C. and A. Perrott, 1978. Date of deglacierisation of Mount Elgon. *Nature* 273: 49.

Hamilton, A. C., D. Taylor and J. C. Vogel, 1989. Neolithic forest clearance at Ahkagyezi, western Uganda. In *Quarternary and Environmental Research on East African Mountains*, edited by Mahaney, W. C., p. 435-463. Rotterdam, A. A. Balkema.

Hamilton, W. D., 1964. The genetical evolution of social behaviour. *J. Theor. Biol.* 7: 1-52.

Harcourt, A. H., 1987. Dominance and fertility among female primates. *J. Zool. Soc., Lond.* 213: 471-487.

Harcourt, A. H., 1996. Is the gorilla a threatened species? How should we judge? *Biol. Cons.* 75: 165-176.

Harpending, H. C., S. T. Sherry, A. R. Rogers and M. Stoneking, 1993. The genetic structure of ancient human populations. *Current Anthro.* 34: 483-496.

Harris, L. D., 1984. *The Fragmented Forest: Island Biogeography Theory and the Preservation of Biotic Diversity*. Chicago, University of Chicago Press.

Harris, R. B. and F. W. Allendorf, 1989. Genetically effective population size of large mammals: an assessment of estimators. *Cons. Biol.* 3: 181-191.

Hart, J., T. Hart and S. C. Thomas, 1986. The Ituri Forest of Zaïre: primate diversity and prospects for conservation. *Primate Cons.* 7: 42-44.

Harvey, P., R. D. Martin and T. H. Clutton-Brock, 1987. Life histories in comparative perspective. In Primate Societies, edited by Smuts, B. B., D. L. Cheney, R. M. Seyfarth, R. W. Wrangham and T. T. Struhsaker, p. 179-196. Chicago, The University of Chicago Press.

Hart, T., J. Hart and P. G. Murphy, 1989. Monodominant and species-rich forests of the humid tropics: causes for their co-occurrence. *Am. Nat.* 133: 613-633.

Hasegawa, T. and M. Hiraiwa-Hasegawa, 1983. Opportunistic and restrictive matings among wild chimpanzees in the Mahale Mountais, Tanzania. *J. Ethol.* 1: 75-85.

Hasegawa, T. and M. Hiraiwa-Hasegawa, 1990. Sperm competition and mating behavior. In *The Chimpanzees of the Mahale Mountains: Sexual and Life-history Strategies*, edited by Nishida, T., p. 115-132. Tokyo, University of Tokyo Press.

Hays, J. D., J. Imbrie and N. J. Shackleton, 1976. Variations in Earth's orbit: pacemaker of the ice ages. *Science* 194: 1121-1132.

Hemelrijk, C. K., 1990a. A matrix partial correlation test used in investigations of reciprocity and other social interaction patterns at group level. *J. Theor. Biol.* 143: 405-420.

Hemelrijk, C. K., 1990b. Models of, and tests for, reciprocity, unidirectionality and other social interaction patterns at a group level. *Anim. Behav.* 39: 1013-1029.

Hennig, W., 1966. *Phylogenetic Systematics*. Urbana, University of Illinois.

Hershkovitz, P., 1969. The recent mammals of the Neotropical Region: a zoogeographic and ecological review. *Quart. Rev. Biol.* 44: 1-70.

Hill, W. C. O., 1969. The nomenclature, taxonomy and distribution of the chimpanzee. In *The Chimpanzee*, edited by Bourne, G. H., p. 22-49. Basel, Karger.

Hillis, D. M., J. J. Bull, M. E. White, M. R. Badgett and I. J. Molineux, 1992. Experimental phylogenetics: generation of a known phylogeny. *Science* 255: 589-592.

Hobbs, R. J., 1992. The role of corridors in conservation: solution or bandwagon? *Tree* 7: 389-392.

Holder, M. and K. Hunt, 1994. *Evidence for intercommunity violence in Kibale Forest chimpanzees*. Unpublished report.

Hope, A. C. A., 1968. A simplified Monte Carlo significance test procedure. *J. Roy. Stat. Soc., Series B.* 30: 582-598.

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. *J. Mol. Evol.* 35: 32-43.

Höss, M., M. Kohn and S. Pääbo, 1992. Excrement analysis by PCR. *Nature* 359: 199.

Howard, P. C., 1986. *Agricultural Encroachment in the Semliki Forest Reserve, Western Uganda.* Report to World Wildlife Fund, New York Zoological Society.

Howard, P. C., 1991. *Nature Conservation in Uganda's Tropical Forest Reserves.* Gland, Switzerland and Cambridge, UK, IUCN.

Hu, G. and W. G. Thilly, 1994. Evolutionary trail of the mitochondrial genome as based on human 16s rDNA pseudogenes. *Gene* 147: 197-204.

Humphries, C. J., P. Y. Ladiges, M. Roos and M. Zandee, 1988. Cladistic biogeography. In *Analytical Biogeography*, edited by Myers, A. A. and P. S. Giller, p. 371-404. London, Chapman & Hall.

Hyder, S. M., C. Hu, D. S. Needleman, Y. Sonoda, X. Wang and V. V. Baker, 1994. Improved accuracy in direct automated DNA sequencing of small PCR products by optimizing the template concentration. *BioTechniques* 17: 478-482.

Imbrie, J., 1985. A theoretical framework for the Pleistocene ice ages. *J. Geol. Soc., Lond.* 142: 417-432.

Imbrie, J., E. A. Boyle, S. C. Clemens, A. Duffy, W. R. Howard, G. Kukla, J. Kutzbach, D. G. Martinson, A. AcIntyre, A. C. Mix, B. Molfino, J. J. Morley, L. C. Peterson, N. G. Pisias, W. L. Prell, M. E. Raymo, N. J. Shackleton and J. R. Toggweiler, 1992. On the structure and origin of major glaciation cycles 1. Linear responses to Milankovitch forcing. *Paleoceanography* 7: 701-738.

Imbrie, J. and J. Z. Imbrie, 1980. Modeling the climatic response to orbital variations. *Science* 207: 943-953.

Innis, M. A., D. H. Gelfand, J. J. Sninsky and T. J. White, Ed., 1990. *PCR Protocols: A Guide to Methods and Applications.* San Diego, Academic Press.

Inoue, K. and T. Kawahara, 1990. Allozyme differentiation and genetic structure in island and mainland Japanese populations of *Campanula punctata* (Campanulaceae). *Am. J. Botany* 77: 1440-1448.

Inoue, M., F. Mitsunaga, H. Ohsawa, A. Takenaka, Y. Sugiyama, A. G. Soumah and O. Takenaka, 1992. Paternity testing in captive Japanese macaques (*Macaca fuscata*) using DNA fingerprinting. In *Paternity in Primates: Genetic Tests and Theories*, edited by Martin, R. D., A. F. Dixson and E. J. Wickings, p. 131-140. Basel, Karger.

Inoue, M., A. Takenaka, S. Tanaka, R. Kominami and O. Takenaka, 1990. Paternity discrimination in a Japanese macaque group by DNA fingerprinting. *Primates* 31: 563-570.

Irwin, D. M., T. D. Kocher and A. C. Wilson, 1991. Evolution of the Cytochrome *b* gene of mammals. *J. Mol. Evol.* 32: 128-144.

Isabirye-Basuta, G., 1991. Nesting behaviour of chimpanzees in Kibale Forest, Uganda. Unpublished Manuscript.

Jacobshagen, B., 1974. Morphometric studies in the taxonomy of the orang-utan (*Pongo pygmaeus*, L. 1760). *Folia Primatol.* 32: 29-34.

Jasper, J. P. and J. M. Hayes, 1990. A carbon isotope record of CO_2 levels during the late Quaternary. *Nature* 347: 462-464.

Jeffreys, A. J., V. Wilson, R. Neumann and J. Keyte, 1988. Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Res.* 16: 10953-10971.

Jeffreys, A. J., V. Wilson and S. L. Thein, 1985. Hypervariable "minisatellite" regions in human DNA. *Nature* 314: 67-73.

Jin, L. and M. Nei, 1990. Limitations of the evolutionary parsimony method of phylogenetic analysis. *Mol. Biol. Evol.* 7: 82-102.

Jolly, D. and R. Bonnefille, 1991. Diagramme pollinique d'un sondage Holocene de la Kuruyange (Burundi, Afrique Centrale). *Palaeoecology of Africa* 22: 265-274.

Kamimura, N., S. Ishii, M. Liandong and J. W. Shay, 1989. Three separate mitochondrial DNA sequences are contiguous in human genomic DNA. *J. Mol. Biol.* 210: 703-707.

Kano, T., 1972. Distribution and adaptation of the chimpanzee on the eastern shore of Lake Tanganyika. *Kyoto University African Studies Monographs* 7: 37-129.

Kendall, R. L., 1969. An ecological history of the Lake Victoria basin. *Ecological Monographs* 39: 121-176.

Kennett, J. P. and R. C. Thunell, 1975. Global increase in Quaternary explosive volcanism. *Science* 187: 497-503.

Kimura, M., 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 61: 893-903.

Kimura, M. and J. F. Crow, 1963. The measurement of effective population number. *Evolution* 17: 279-288.

Kimura, M. and G. H. Weiss, 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49: 561-576.

Kingdon, J., 1981. Where have the colonists come from? A zoogeographical examination of some mammalian isolates in eastern Africa. *African J. Ecol.* 19: 115-124.

Kingdon, J., 1989. *Island Africa: The Evolution of Africa's Rare Animals and Plants.* Princeton, Princeton University Press.

Kisubi, A. and P. Howard, 1990. *Use of 'nest' counts in evaluating chimpanzee habitat preferences in the forests of western Uganda.* Unpublished report.

Kocher, T. D. and A. C. Wilson, 1991. Sequence evolution of mitochondrial DNA in humans and chimpanzees. In *Evolution of Life: Fossils, Nucleotides and Culture*, edited

by p. 391-413. Tokyo, Springer.

Kortlandt, A., 1972. *New Perspectives in Ape and Human Evolution.* Amsterdam, Stichting von voor Psychobiologie.

Kortlandt, A., 1983. Marginal habitats of chimpanzees. *J. Human Evol.* 12: 231-278.

Kusukawa, N., T. Uemori, K. Asada and I. Kato, 1990. Rapid and reliable protocol for direct sequencing of material amplified by the polymerase chain reaction. *BioTechniques* 66-72.

Kwok, S., 1990. Procedures to minimize PCR-product carry-over. In *PCR Protocols: A Guide to Methods and Applications*, edited by Innis, M. A., D. H. Gelfand, J. J. Sninsky and T. J. White, p. 142-145. San Diego, Academic Press.

Laikre, L. and N. Ryman, 1991. Inbreeding depression in a captive wolf (*Canis lupus*) population. *Cons. Biol.* 5: 33-40.

Lamb, T. and J. C. Avise, 1992. Molecular and population genetics aspects of mitochondrial DNA variability in the diamondback terrapin, *Malaclemys terrapin. J. Heredity* 83: 262-269.

Lande, R., 1988. Genetics and demography in biological conservation. *Science* 241: 1455-1460.

Lande, R., 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *Am. Nat.* 141: 911-927.

Lande, R. and G. F. Barrowclough, 1987. Effective population size, genetic variation, and their use in population management. In *Viable Populations for Conservation*, edited by Soulé, M. E., p. 87-123. Cambridge, Cambridge University Press.

Langdale-Brown, I., H. A. Osmaston and J. G. Wilson, 1964. *The Vegetation of Uganda and its Bearing on Land Use.* Entebbe, The Government Printer.

Latter, B. D. H., 1980. Genetic differences within and between populations of the major human subgroups. *Am. Nat.* 116: 220-237.

Ledig, F. T., 1986. Heterozygosity, heterosis and fitness in outbreeding plants. In *Conservation Biology: The Science of Scarcity and Diversity*, edited by Soulé, M. E., p. 77-104. Sunderland, Sinauer Associates, Inc.

Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74: 1659-1673.

Legendre, P. and M. J. Fortin, 1989. Spatial pattern and ecological analysis. *Vegetatio* 80: 107-138.

Legendre, P. and A. Vaudour, 1991. *The R Package: Multidimensional Analysis, Spatial Analysis.* Version 3. Université de Montréal, Département de Sciences Biologiques.

Lehman, N. and R. K. Wayne, 1991. Analysis of coyote mitochondrial DNA genotype frequencies: estimation of the effective number of alleles. *Genetics* 128: 405-416.

Leuenberger, M., U. Siegenthaler and C. C. Langway, 1992. Carbon isotope composition of atmospheric CO_2 during the last ice age from an Antarctic ice core. *Nature* 357: 488-490.

Leung, D. W., E. Chen and D. V. Goeddel, 1989. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* 1: 11-15.

Lewontin, R. C., 1972. The apportionment of human diversity. *Evol. Biol.* 6: 381-398.

Lezine, A.-M. and J. Casanova, 1991. Correlated oceanic and continental records demonstrate past climate and hydrology of North Africa (0-140 ka). *Geology* 19: 307-310.

Li, W.-H. and D. Graur, 1991. *Fundamentals of Molecular Evolution.* Sunderland, Sinauer Associates.

Lillefors, H. W., 1967. On the Kolmolgorov-Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* 64: 399-402.

Liu, H.-s., 1992. Frequency variations of the Earth's obliquity and the 100-kyr ice-age cycles. *Nature* 358: 397-399.

Livingstone, D. A., 1962. Age of deglaciation in the Ruwenzori range, Uganda. *Nature* 194: 859-860.

Livingstone, D. A., 1975. Late Quaternary climatic change in Africa. *Ann. Rev. Ecol. Syst.* 6: 249-280.

Long, J. C., 1986. The allelic correlation structure of Gainj- and Kalam-speaking people. I. The estimation and interpretation of Wright's *F*-statistics. *Genetics* 112: 629-647.

Lonnberg, E., 1929. The development and distribution of the African fauna in connection with and depending upon climatic changes. *Ark. Zool.* 21A: 1-33.

Lundberg, J. G., 1972. Wagner networks and ancestors. *Sys. Zool.* 21: 398-413.

MacArthur, R. H. and E. O. Wilson, 1967. *The Theory of Island Biogeography.* Princeton, Princeton University Press.

MacNeil, D. and C. Strobeck, 1987. Evolutionary relationships among colonies of Columbian ground squirrels as shown by mitochondrial DNA. *Evolution* 41: 873-881.

Maddison, D. R., M. Ruvolo and D. L. Swofford, 1992. Geographic origins of human mitochondrial DNA: phylogenetic evidence from control region sequences. *Syst. Biol.* 41: 111-124.

Maddison, W. P. and D. R. Maddison, 1992. *MacClade: Analysis of Phylogeny and Character Evolution.* Version 3.0. Sunderland, Massachusetts, Sinauer Associates.

Major, J., 1988. Endemism: a botanical perspective. In *Analytical Biogeography*, edited by Myers, A. A. and P. S. Giller, p. 117-148. London, Chapman & Hall.

Malenky, R. K. and R. W. Wrangham, 1994. Quantitative comparison of terrestrial herbaceous food consumption by *Pan paniscus* in the Lomako Forest, Zaire, and *Pan troglodytes* in the Kibale Forest, Uganda. *Am. J. Primatol.* 32: 1-12.

Maley, J., 1987. Fragmentation de la forêt dense humide Africaine et extension des biotropes montagnardes au Quaternaire récent: Nouvelles données polliniques et chronologiques. Implications paleoclimatiques et biogeographiques. *Palaeoecology of Africa* 18: 307-333.

Maley, J., D. A. Livingstone, P. Giresse, P. Brenac, G. Kling, C. Stager, N. Thouveny, K. Kelts, M. Haag, M. Fournier, Y. Bandet, D. Williamson and A. Zogning, 1991. West Cameroon Quaternary lacustrine deposits: preliminary results. *J. Afr. Earth Sci.* 12: 147-157.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27: 209-220.

Marjoram, P. and P. Donnelly, 1994. Pairwise comparisons of mirochondrial DNA sequences in subdivided populations and implications for early human evolution. *Genetics* 136: 673-683.

Martin, R. D., A. F. Dixson and E. J. Wickings, Ed., 1992. *Paternity in Primates: Genetic Tests and Theories*. Basel, Karger.

Mayr, E., 1954. Changes of genetic environment and evolution. In *Evolution as a Process*, edited by Huxley, J., A. C. Hardy and E. B. Ford, p. 157-180. London, Allen and Unwin.

Mayr, E., 1963. *Animal Species and Evolution*. Cambridge, Belknap Press of Harvard University Press.

Mayr, E. and R. J. O'Hara, 1986. The biogeographic evidence supporting the Pleistocene forest refuge hypothesis. *Evolution* 41: 55-67.

McClenaghan, L. R., J. Berger and H. Truesdale, 1990. Founding lineages and genetic variability in plains bison (*Bison bison*) from Badlands National Park, South Dakota. *Cons. Biol.* 4: 285-289.

McCoy, E. D. and K. L. Heck, 1983. Centers of origin revisited. *Paleobiolgy* 9: 17-19.

McGrew, W. C., 1985. The chimpanzee and the oil palm: patterns of culture. *Social Biology and Human Affairs* 50: 7-23.

McGrew, W. C., 1992. *Chimpanzee Material Culture*. Cambridge, Cambridge University Press.

McGrew, W. C., 1994. Tools compared: The material of culture. In *Chimpanzee Cultures*, edited by Wrangham, R. W., W. C. McGrew, F. B. M. de Waal and P. G. Heltne, p. 25-40. Cambridge, Harvard University Press.

McGrew, W. C., P. J. Baldwin and C. E. G. Tutin, 1981. Chimpanzees in a hot, dry and open habitat: Mt. Assirik, Senegal, West Africa. *J. Hum. Evol.* 10: 227-244.

Medori, R., H. J. Tritschler and P. Gambetti, 1992. Production of single stranded DNA for sequencing: an alternative approach. *BioTechniques* 12: 346-350.

Merriwether, D. A., A. G. Clark, S. W. Ballinger, T. G. Schurr, H. Soodyall, T. Jenkins, S. T. Sherry and D. C. Wallace, 1991. The structure of human mitochondrial DNA variation. *Mol. Evol.* 33: 543-555.

Milankovitch, M., 1930. Mathematische Klimalehre und astronomische Theorie der Klimaschwankungen. In *Handbuch der Klimatologie*, edited by Köppen, W. and R. Geiger, p. Berlin, Borntraeger.

Moeyersons, J. and E. Roche, 1982. Past and present environment. In *The Archaeology of Central Africa*, edited by Van Noten, F., p. 15-26. Graz/Austria, Akademische Druck u. Verlagsanstalt.

Moore, J., 1992a. "Savanna" chimpanzees. In *Topics in Primatology: Human Origins*, edited by Nishida, T., W. C. McGrew, P. Marler, M. Pickford and F. B. M. de Waal, p. 99-118. Tokyo, University of Tokyo Press.

Moore, J., 1992b. Chimpanzees, conservation, and hair. *Occ. Papers of the Museum of Anthropology, CSU Fullerton* 6: 25-38.

Moore, W. S., J. H. Graham and J. T. Price, 1991. Mitochondrial DNA variation in the Northern Flicker (*Colaptes auratus*, Aves). *Mol. Biol. Evol.* 8: 327-344.

Moran, P. A. P., 1950. Notes on continuous stochastic phenomena. *Biometrika* 37: 17-23.

Moreau, R. E., 1963. Vicissitudes of the African biomes in the late Pleistocene. *Proc. Zool. Soc. Lond.* 141: 395-422.

Moreau, R. E., 1965. Climatic changes and the distribution of forest vertebrates in West Africa. *J. Zool. Soc. Lond.* 185: 39-61.

Moreau, R. E., 1966. *The Bird Faunas of Africa and its Islands*. London, Academic Press.

Morin, P. A., 1993. Reproductive strategies in chimpanzees. *Yearbook Phys. Anthro.* 36: 179-212.

Morin, P. A., J. J. Moore, R. Chakraborty, L. Jin, J. Goodall and D. S. Woodriff, 1994a. Kin selection, social structure, gene flow and the evolution of chimpanzees. *Science* 265: 1145-1332.

Morin, P. A., J. Wallis, J. J. Moore and D. S. Woodruff, 1994b. Paternity exclusion in a community of wild chimpanzees using hypervariable simple sequence repeats. *Mol. Ecol.* 3: 469-478.

Morin, P. A. and D. S. Woodruff, 1992. Paternity exclusion using multiple hypervariable microsatellite loci amplified from nuclear DNA of hair cells. In *Paternity in Primates: Genetic Tests and Theories*, edited by Martin, R. D., A. F. Dixson and E. J. Wickings, p.

63-81. Basel, Karger.

Moritz, C., 1991. The origin and evolution of pathogenesis in *Heteronotia binoei* (Gekkonidae): evidence for recent and localized origins of widespread clones. *Genetics* 129: 211-219.

Morrison, M. E. S., 1961. Pollen analysis in Uganda. *Nature* 190: 483-486.

Morrone, J. J. and J. M. Carpenter, 1994. In search of a method for cladistic biogeography: an empirical comparison of component analysis, brooks parsimony analysis and three-area statements. *Cladistics* 10: 99-153.

Muroyama, Y. and Y. Sugiyama, 1994. Grooming relationships in two species of chimpanzees. In *Chimpanzee Cultures*, edited by Wrangham, R. W., W. C. McGrew, F. B. M. de Waal and P. G. Heltne, p. 169-180. Cambridge, Harvard University Press.

Murray, J., O. C. Stine and M. S. Johnson, 1991. The evolution of mitochondrial DNA in *Partula*. *Heredity* 66: 93-104.

Myers, A. A. and P. S. Giller, 1988. Process, pattern and scale in biogeography. In *Analytical Biogeography*, edited by Myers, A. A. and P. S. Giller, p. 3-12. London, Chapman & Hall.

Neftel, A., H. Oeschger, T. Staffelbach and B. Stauffer, 1988. CO2 record in the Byrd ice core 50,000-5,000 years BP. *Nature* 331: 609-611.

Nei, M., 1972. Genetic distance between populations. *Am. Nat.* 106: 238-292.

Nei, M., 1973. Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci., USA*. 70: 3321-3323.

Nei, M., 1977. *F*-statistics and analysis of gene diversity in subdivided populations. *Ann. Hum. Gen.* 41: 225-233.

Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583-590.

Nei, M., F. Tajima and Y. Tateno, 1983. Accuracy of estimated phylogenetic trees from molecular data. *J. Mol. Evol.* 19: 153-170.

Nei, M., 1987. *Molecular Evolutionary Genetics*. New York, Columbia University Press.

Nei, M. and D. Graur, 1984. Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.* 17: 73-118.

Nei, M., T. Maruyama and R. Chakraborty, 1975. The bottleneck effect and genetic variability in populations. *Evolution* 29: 1-10.

Nei, M. and A. K. Roychoudhury, 1974. Genetic variation within and between the major races of man, Caucasoids, Negroids and Mongoloids. *Am. J. Hum. Genet.* 26: 421-443.

Nei, M., F. Tajima and Y. Tateno, 1983. Accuracy of estimated phylogenetic trees from

molecular data. *J. Mol. Evol.* 19: 153-170.

Nelson, G., 1981. *Systematics and biogeography/cladistics and vicariance.* New York, Columbia University Press.

Newmark, W. D., 1991. Tropical forest fragmentation and the local extinction of understory birds in the eastern Usambara Mountains, Tanzania. *Cons. Biol.* 5: 67-78.

Nishida, T., 1968. The social group of wild chimpanzees in the Mahale mountains. *Primates* 9: 167-224.

Nishida, T., 1979. The social structure of chimpanzees of the Mahale mountains. In *The Great Apes*, edited by Hamburg, D. A. and E. R. McCown, p. 73-122. Menolo Park, Benjamin/Cummings.

Nishida, T., Ed., 1990. *The Chimpanzees of the Mahale Mountains: Sexual and Life History Strategies.* Tokyo, University of Tokyo Press.

Nishida, T., 1994. Review of recent findings on Mahale chimpanzees. In *Chimpanzee Cultures*, edited by Wrangham, R. W., W. C. McGrew, F. B. M. de Waal and P. G. Heltne, p. 373-396. Cambridge, Harvard University Press.

Nishida, T. and M. Hiraiwa-Hasegawa, 1986. Chimpanzees and bonobos: cooperative relationships among males. In *Primate Societies*, edited by Smuts, B. B., D. L. Cheney, R. M. Seyfarth, R. W. Wrangham and T. T. Struhsaker, p. 165-177. Chicago, University of Chicago Press.

Nissen, H. W., 1931. A field study of the chimpanzee. Observations of chimpanzee behavior and environment in Western French Guinea. *Comp. Psychol. Monogr.* 8: 1-122.

Nomiyama, H., M. Fukuda, S. Wakesugi, T. Tsuzuki and K. Shimada, 1985. Molecular structures of mitochondrial-DNA-like sequences in human nuclear DNA. *Nucleic Acids Res.* 13: 1649-1658.

O'Brien, S. J., J. S. Martenson, M. A. Eichelberger, E. T. Thorne and F. Wright, 1989. Genetic variation and molecular systematics of the black-footed ferret. In *Conservation Biology and the Black-Footed Ferret*, edited by Seal, U. S., E. T. Thorne, M. A. Bogan and S. H. Anderson, p. 21-33. New Haven, Yale University Press.

O'Brien, S. J., D. E. Wildt, M. Bush, T. M. Caro, C. FitzGibbon, I. Aggundey and R. E. Leakey, 1987. East African cheetahs: evidence for two population bottlenecks? *Proc. Nat. Acad. Sci., USA.* 84: 508-511.

O'Brien, S. J., D. E. Wildt, D. Goldman, D. R. Merril and M. Bush, 1983. The cheetah is depauperate in genetic variation. *Science* 221: 459-462.

Oden, N. L., 1984. Assessing the significance of a spatial correlogram. *Geog. Anal.* 16: 1-16.

Osmaston, H. A., 1967. The sequence of glaciations in the Ruwenzori and their correlation with glaciations of other mountains in East Africa and Ethiopia. *Palaeoecology of Africa* 2: 26-28.

Pääbo, S., 1990. Amplifying ancient DNA. In *PCR Protocols: A Guide to Methods and Applications*, edited by Innis, M. A., D. H. Gelfand, J. J. Sninsky and T. J. White, p. 159-166. San Diego, Academic Press.

Pääbo, S., J. A. Gifford and A. C. Wilson, 1988. Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids Res.* 16: 9775-9787.

Packer, C., A. E. Pusey, H. Rowley, D. A. Gilbert, J. S. Martinson and S. J. O'Brien, 1991. Case study of a population bottleneck: lions of the Ngorogoro Crater. *Cons. Biol.* 5: 1-12.

Partridge, T. C., 1993. Warming phases in Southern Africa during the last 150,000 years: an overview. *Palaeogeog., Palaeoclimatol., Palaeoecol.* 101: 237-244.

Peet, R. K., 1974. Measurement of species diversity. *Ann. Rev. Ecol. Syst.* 5: 285-307.

Penny, D., 1982. Towards a basis for classification: the incompleteness of distance measures, incompatibility analysis and phenetic classification. *J. Theor. Biol.* 96: 129-142.

Phillipson, D. W., 1977. *The Later Prehistory of Eastern and Southern Africa.* London, Heinemann.

Pilbeam, D., 1996. Genetic and morphological records of the Hominoidea and hominid origins: a synthesis. *Mol. Phyl. Evol.* 5: 155-168.

Plante, Y., P. T. Boag and B. N. White, 1989. Microgeographic variation in the mitochondrial DNA of meadow voles (*Microtus pennsylvanicus*) in relation to population density. *Evolution* 43: 1522-1537.

Platnick, N. I. and G. Nelson, 1984. Composite areas in vicariance biogeography. *Sys. Zool.* 33: 328-335.

Prinsloo, P. and T. J. Robinson, 1992. Geographic mitochondrial DNA variation in the rock hyrax, *Procavia capensis. Mol. Biol. Evol.* 9: 447-456.

Pusey, A. E., 1980. Inbreeding avoidance in chimpanzees. *Anim. Behav.* 28: 543-552.

Ralls, K. and J. Ballou, 1982. Effect of inbreeding on infant mortality in captive primates. *Int. J. Primatol.* 3: 491-505.

Ralls, K., K. Brugger and J. Ballou, 1979. Inbreeding and juvenile mortality in small populations of ungulates. *Science* 206: 1101-1103.

Ralls, K., P. H. Harvey and A. M. Lyles, 1986. Inbreeding in natural populations of birds and mammals. In *Conservation Biology: The Science of Scarcity and Diversity*, edited by Soulé, M. E., p. 35-56. Sunderland, Sinauer Associates, Inc.

Raymo, M. E. and W. F. Ruddiman, 1992. Tectonic forcing of late Cenozoic climate. *Nature* 359: 117-122.

Reynolds, J., B. S. Weir and C. C. Cockerham, 1983. Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics* 105: 767-779.

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

Rice, K. A. and M. J. Donoghue, 1996. Reanalysis of a large *rbcL* dataset: implications for future phylogenetic studies. *Systematic Biology* (In Press).

Riddle, B. R. and R. L. Honeycutt, 1990. Historical biogeography in North American arid regions: an approach using mitochondrial-DNA phylogeny in grasshopper mice (genus *Onychomys*). *Evolution* 44: 1-15.

Rodgers, W. A., C. F. Owen and K. M. Homewood, 1982. Biogeography of East African forest mammals. *J. Biogeog.* 9: 41-54.

Rodman, P. S. and H. M. McHenry, 1980. Bioenergetics and the origin of hominoid bipedalism. *Am. J. Phys. Anthro.* 52: 103-106.

Rogers, A., 1992. Error introduced by the infinite sites model. *Mol. Biol. Evol.* 9: 1181-1184.

Rogers, A. R., 1995a. Genetic evidence for a pleistocene population explosion. *Evolution* 49: 608-615.

Rogers, A. R., 1995b. *MISMATCH: computer programs for analysis of mismatch distributions.* Version 2b. Salt Lake City.

Rogers, A. R., 1995c. Population structure and modern human origins. In *Progress in Population Genetics and Human Evolution*, edited by Donnelly, P. J. and S. Tavaré. New York, Springer-Verlag.

Rogers, A. R. and H. Harpending, 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* 9: 552-569.

Rogers, A. R. and L. B. Jorde, 1995. Genetic evidence on modern human origins. *Hum. Biol.* 67: 1-36.

Rogers, J., 1992. Nuclear DNA polymorphism in Hominoids and Cercopithecoids: applications to paternity testing. In *Paternity in Primates: Genetic Tests and Theories*, edited by Martin, R. D., A. F. Dixson and E. J. Wickings, p. 82-95. Basel, Karger.

Rosen, B. R., 1984. Reef coral biogeography and climate through the late Cainozoic: just islands in the sun or a critical pattern of islands? In *Fossils and Climate*, edited by Brenchley, P. J., p. 201-262. London, Geol. J. Special Issue.

Rosen, B. R., 1985. Geological hierarchies and biogeographic congruence in the Caribbean. *Ann. Missouri Botanical Garden* 72: 636-659.

Rosen, B. R., 1988. From fossils to Earth history: applied historical biogeography. In *Analytical Biogeography*, edited by Myers, A. A. and P. S. Giller, p. 437-481. London, Chapman & Hall.

Rosen, B. R. and A. B. Smith, 1988. Tectonics from fossils? Analysis of reef coral and sea urchin distributions from late Cretaceous to Recent, using a new method. In *Gondwana and Tethys*, edited by Audley-Charles, M. G. and A. Hallam, p. 275-306. Lonon, Geological Society of London Special Publicaion.

Ruano, G., L. Rogers, A. C. Ferguson-Smith and K. K. Kidd, 1992. DNA sequence polymorphisms within hominoid species exceeds the number of phylogenetically informative characters for a HOX2 locus. *Mol. Biol. Evol.* 9: 575-586.

Ruvolo, M., 1996a. Molecular phylogeny of the Hominoids: application of the likelihood ratio method. *Mol. Biol. Evol.* (submitted).

Ruvolo, M., 1996b. A new approach to studying modern human origins: hypothesis testing with coalescence time distributions. *Molecular Phylogeny and Evolution* 5: 202-219.

Ruvolo, M., T. R. Disotell, M. W. Allard, W. M. Brown and R. L. Honeycutt, 1991. Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. *Proc. Nat. Acad. Sci., USA* 88: 1570-1574.

Ruvolo, M., D. Pan, S. Zehr, T. Goldberg, T. Disotell and M. von Dornum, 1994. Gene trees and hominoid phylogeny. *Proc. Nat. Acad. Sci., USA* 91: 8900-8904.

Ruvolo, M., S. Zehr, M. von Dornum, D. Pan, B. Chang and J. Lin, 1993. Mitochondrial COII sequences and modern human origins. *Mol. Biol. Evol.* 10: 1115-1135.

Sage, R. D. and J. O. Wolff, 1986. Pleistocene glaciations, fluctuating ranges, and low genetic variability in a large mammal (*Ovis dalli*). *Evolution* 40: 1092-1095.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. Erlich, 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487-491.

Saiki, R. K., S. J. Scharf, F. Faloona, K. B. Mullis, G. T. Horu, H. A. Erlich and N. Arnheim, 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350-1354.

Saitou, N. and M. Nei, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, Cold Spring Harbor Laboratory Press.

Sarnthein, M., 1978. Sand deserts during glacial maximum and climatic optimum. *Nature* 272: 43-46.

Saunders, D. A., 1991. Biological consequences of ecosystem fragmentation: a review. *Cons. Biol.* 5: 18-32.

Saunders, D. A. and R. J. Hobbs, Ed., 1991. *The Role of Corridors.* New South Wales, Surrey Beatty, Chipping Norton.

Schonewald-Cox, C. M., S. M. Chambers, B. MacBryde and W. L. Thomas, Ed., 1983. *Genetics and Conservation: A Reference for Managing Wild Animal and Plant Populations*. Menlo Park, The Benjamin/Cummings Publishing Company, Inc.

Schwartz, E., 1934. On the local races of the chimpanzee. *Ann. Mag. Nat. Hist. Lond.* 13: 576-583.

Senner, J. W., 1980. Inbreeding depression and the survival of zoo populations. In *Conservation Biology: An Evolutionary-Ecological Perspective*, edited by Soulé, M. E. and B. A. Wilcox, p. 209-224. Sunderland, Sinauer.

Sept, J. M., 1992. Was there no place like home? A new perspective on early hominid archaeological sites from the mapping of chimpanzee nests. *Current Anthro.* 33: 187-208.

Seutin, G., J. Brawn, R. E. Ricklefs and E. Bermingham, 1993. Genetic divergence among populations of a tropical passerine, the streaked saltator (*Saltator albicollis*). *Auk* 110: 117-126.

Seutin, G., N. K. Klein, R. E. Ricklefs and E. Bermingham, 1994. Historical biogeography of the bananaquit (*Coereba flaveola*) in the Caribbean region: a mitochondrial DNA assesment. *Evolution* 48: 1041-1061.

Shackleton, N. J. and N. D. Opdyke, 1973. Oxygen isotope and paleomagnetic stratigraphy of equatorial Pacific core V28-238: oxygen isotope temperatures and ice volumes on a 10^5 and 10^6 year scale. *Quat. Res.* 3: 39-55.

Shackleton, N. J., R. G. West and D. Q. Bowen, Ed., 1988. *The Past Three Million Years: Evolution of Climatic Variability in the North Atlantic Region*. London, The Royal Society.

Shea, B. T. and H. J. Coolidge, 1988. Craniometric differentiation and systematics in the genus *Pan*. *J. Hum. Evol.* 17: 671-685.

Sherry, S. T., A. R. Rogers, H. Harpending, H. Soodyall, T. Jenkins and M. Stoneking, 1994. Mismatch distributions of mtDNA reveal recent human population expansions. *Hum. Biol.* 66: 761-776.

Sherwin, W. B., N. D. Murray, J. A. M. Graves and P. R. Brown, 1991. Measurement of genetic variation in endangered populations: bandicoots (Marsupiala: Peramelidae) as an example. *Cons. Biol.* 5: 105-108.

Shields, G. F. and A. C. Wilson, 1987. Subspecies of the Canada Goose (*Branta canadensis*) have distinct mitochondrial DNAs. *Evolution* 41: 662-666.

Sibley, C. G. and J. E. Ahlquist, 1987. The phylogeny of the hominoid primates as indicated by DNA-DNA hybridization. *J. Mol. Evol.* 20: 2-15.

Silk, J. B., 1987. Social behavior in evolutionary perspective. In *Primate Societies*, edited by Smuts, B. B., D. L. Cheney, R. M. Seyfarth, R. W. Wrangham and T. T. Struhsaker, p. 318-329. Chicago, The University of Chicago Press.

Simberloff, D., J. A. Farr, J. Cox and D. Mehlman, 1992. Movement corridors:

conservation bargains or poor investments? *Cons. Biol.* 6: 493-504.

Skinner, B. J. and S. C. Parker, 1987. *Physical Geology*. New York, John Wiley and Sons.

Slatkin, M. and R. R. Hudson, 1991. Pairwise comparison of mitochondrial DNA sequences in stable and exponenetially growing populations. *Genetics* 129: 555-562.

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

Smith, D. G., 1981. The association between rank and reproductive success of male rhesus monkeys. *Am. J. Primatol.* 1: 83-90.

Smouse, P. E., J. C. Long and R. R. Sokal, 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Sys. Zool.* 35: 627-632.

Sober, E., 1988. *Reconstructing the Past: Parsimony, Evolution and Inference.* Cambridge, MIT Press.

Sokal, R. R., R. M. Harding and N. L. Oden, 1989. Spatial patterns of human gene frequencies in Europe. *Am. J. Phys. Anthro.* 80: 267-294.

Sokal, R. R. and P. Menozzi, 1982. Spatial autocorrelation of HLA frequencies in Europe support demic diffusion of early farmers. *Am. Nat.* 119: 1-17.

Sokal, R. R. and C. D. Michener, 1958. A statistical method for evaluating systematic relationships. *U. Kansas Sci. Bull.* 28: 1409-1438.

Sokal, R. R. and N. L. Oden, 1978a. Spatial autocorrelation in biology 1: methodology. *Biol. J. Linn. Soc.* 10: 199-228.

Sokal, R. R. and N. L. Oden, 1978b. Spatial autocorrelation in biology 2: some biological implications and four applications of evolutionary and ecological interest. *Biol. J. Linn. Soc.* 10: 229-249.

Sounders, N. C., L. G. Kessler and J. C. Avise, 1986. Genetic variation and geographic differentiation in mitochondrial DNA of the horseshoe crab, *Limulus polyphemus*. *Genetics* 112: 613-627.

Sowunmi, M. A., 1991. Late Quaternary environments in equatorial Africa: palynological evidence. *Palaeoecology of Africa* 22: 213-238.

Stewart, C.-B., 1995. Active ancestral molecules. *Nature* 374: 12-13.

Stoneking, M., L. B. Jorde, K. Bhatia and A. C. Wilson, 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124: 717-733.

Street, F. A. and A. T. Grove, 1979. Global maps of lake-level fluctuations since 30,000 yr B.P. *Quat. Res.* 12: 83-118.

Struhsaker, T. T., 1972. Rain-forest conservation in Africa. *Primates* 13: 103-109.

Struhsaker, T. T., 1975. *The Red Colobus Monkey*. Chicago, University of Chicago Press.

Struhsaker, T. T., 1981. Forest and primate conservation in East Africa. *Afr. J. Ecol.* 19: 99-114.

Struhsaker, T. T., 1987. Forestry Issues and Conservation in Uganda. *Biol. Cons.* 39: 209-234.

Sugiyama, Y., 1988. Grooming interactions among adult chimpanzees at Bossou, Guinea, with special reference to social structure. *Int. J. Primatol.* 9: 393-407.

Sugiyama, Y. and J. Koman, 1979. Social structure and dynamics of wild chimpanzees at Bossou, Guinea. *Primates* 20: 323-339.

Swofford, D. L., 1993. *PAUP: Phylogenetic Analysis Using Parsimony*. Version 3.1.1. Illinois Natural History Survey, Champaign.

Swofford, D. L. and G. J. Olsen, 1990. Phylogeny reconstruction. In *Molecular Systematics*, edited by Hillis, D. M. and C. Moritz, p. 411-501. Sunderland, Sinauer.

Taberlet, P. and J. Bouvet, 1992. Bear conservation genetics. *Nature* 358: 197.

Tahara, T., J. P. Kraus and L. E. Rosenberg, 1990. Direct sequencing of PCR amplified genomic DNA by the Maxam-Gilbert method. *BioTechniques* 8: 366-368.

Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.

Tajima, F., 1989. The effect of change in population size on DNA polymorphism. *Genetics* 123: 597-601.

Takahata, N., 1993. Allelic geneaology and human evolution. *Mol. Biol. Evol.* 10: 2-22.

Takahata, N. and S. R. Palumbi, 1985. Extranuclear differentiation and gene flow in the finite island model. *Genetics* 109: 441-457.

Taylor, D. M., 1990. Late Quaternary pollen records from two Ugandan mires: evidence for environmental change in the Rukiga Highlands of southwest Uganda. *Paleogeog., Paleoclimat., Paleoecol.* 80: 283-300.

Taylor, D. M., 1993. Environmental change in montane southwest Uganda: a pollen record for the Holocene from Ahakagyezi Swamp. *The Holocene* 3: 324-332.

Taylor, K. C., G. W. Lamorey, G. A. Doyle, R. B. Alley, P. M. Grootes, P. A. Mayewski, J. W. C. White and L. K. Barlow, 1993. The 'flickering switch' of late Pleistocene climate change. *Nature* 361: 432-436.

Teleki, G., 1989. Population status of wild chimpanzees (*Pan troglodytes*) and threats to survival. In *Understanding Chimpanzees*, edited by Heltne, P. G. and L. A. Marquardt, p. 312-353. Cambridge, Harvard University Press.

Templeton, A. R., 1993. The "Eve" hypothesis: A genetic critique and reanalysis. *Am. Anthropol.* 95: 51-72.

Terborgh, J. and B. Winter, 1980. Some causes of extinction. In *Conservation Biology: An Evolutionary-Ecological Perspective*, edited by Soulé, M. E. and B. A. Wilcox, p. 119-133. Sunderland, Sinauer.

Thomas, S. C., 1991. Population densities and patterns of habitat use among anthropoid primates of the Ituri Forest, Zaïre. *Biotropica* 23: 68-83.

Thompson, L. G., E. Mosley-Thompson, M. E. Davis, P. Lin, K. A. Henderson, J. Cole-Dai, J. F. Bolzan and K. Liu, 1995. Late glacial stage records and Holocene tropical ice core records from Huascarán, Peru. *Science* 269: 46-50.

Tishkoff, S. A., E. Dietzsch, W. Speed, A. J. Pakstis, J. R. Kidd, K. Cheung, B. Bonné-Tamir, A. S. Santacharia-Benerecetti, P. Moral, M. Krings, S. Pääbo, E. Watson, N. Risch, T. Jenkins and K. K. Kidd, 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* 271: 1380-1387.

Tutin, C. E. G., 1979. Mating patterns and reproductive strategies in a community of wild chimpanzees (*Pan troglodytes schweinfurthii*). *Behav. Ecol. Sociobiol.* 6: 29-38.

Tutin, C. E. G., 1980. Reproductive behaviour of wild chimpanzees in the Gombe National Park, Tanzania. *J. Reprod. Fertil. Supp.* 28: 43-57.

Tutin, C. E. G. and M. Fernandez, 1984. Nationwide census of gorilla (*Gorilla g. gorilla*) and Chimpanzee (*Pan t. troglodytes*) populations in Gabon. *Am. J. Primatol.* 6: 313-336.

Uchida, A., 1992. *Intra-species Variation among the Great Apes: Implications for Taxonomy of Fossil Hominoids.* Ph. D. thesis. Harvard University.

Van Campo, E., J. C. Duplessy, W. L. Prell, N. Barratt and R. Sabatier, 1990. Comparison of terrestrial and marine temperature estimates for the past 135 kyr off southeast Africa: a test for GCM simulations of paleoclimate. *Nature* 348: 209-212.

van Hoof, J. A. R. A. M., 1994. Understanding chimpanzee understanding. In *Chimpanzee Cultures*, edited by Wrangham, R. W., W. C. McGrew, F. B. M. de Waal and P. G. Heltne, p. 267-284. Cambridge, Harvard University Press.

Van Neer, W., 1984. Faunal remains from Matupi Cave, an iron age and late stone age site in northeastern Zaïre. *Overdruk uit Academiae Analecta* 46: 59-76.

van Zinderen Bakker, E. M. and J. A. Coetzee, 1972. A re-appraisal of late-Quaternary climatic evidence from tropical Africa. *Palaeoecology of Africa* 7: 151-181.

Verdcourt, B., 1972. The zoogeography of the non-marine Mollusca of East Africa. *J. Conchology* 27: 291-348.

Vigilant, L., R. Pennington, H. Harpending, T. D. Kocher and A. C. Wilson, 1989. Mitochondrial DNA sequences in single hairs from a southern African population. *Proc. Nat. Acad. Sci., USA* 86: 9350-9354.

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.

Vincens, A., 1986. Diagramme pollinique d'un sondage Pleistocene superieur-Holocene du lac Bogoria (Kenya). *Rev. Palaeobotany and Palynology* 47: 169-192.

Vincens, A., 1989. Paleoenvironments du bassin nord-Tanganyika (Zaïre, Burundi, Tanzanie) au cours des 13 derniers mille ans: apport de la palynologie. *Rev. Palaeobotany and Palynology* 61: 69-88.

Vincens, A., 1991. Late Quaternary vegetation history of the south-Tanganyika basin. Climatic implications in south central Africa. *Palaeogeog., Palaeoclimatol., Palaeoecol.* 86: 207-226.

von Koenigswald, G. H. R., 1982. Distribution and evolution of the orang-utan, *Pongo pygmaeus* (Hoppius). In *The Orang Utan: Its Biology and Conservation*, edited by Boer, L. E. M. d. and W. Junk, p. 1-15. Boston, The Hague.

Vrba, E. S., 1985. Ecological and adaptive changes associated with early Hominid evolution. In *Ancestors: The Hard Evidence*, edited by Delson, E., p. 63-71. New York, Alan R. Liss, Inc.

Vrba, E. S., 1989. Late Pliocene climatic events and Hominoid Evolution. In *Evolutionary History of the "Robust" Australopithecines*, edited by Grine, F., p. 405-426. Chicago, Aldine.

Vrba, E. S., 1993. The pulse that produced us. *Nat. Hist.* 102: 47-51.

Vrijenhoek, R. C., 1989. Population genetics and conservation. In *Conservation for the Twenty-first Century*, edited by Western, D. and M. C. Pearl, p. 89-98. New York, Oxford University Press.

Wada, S., T. Kobayashi and K. Numachi, 1991. Genetic variability and differentiation of mitochondrial DNA in Minke whales. In *Genetic Ecology of Whales and Dolphins*, edited by Hoelzel, A. R., p. 203-215. Cambridge, England, Black Bear Press.

Wallis, G. P. and J. W. Arntzen, 1989. Mitochondrial-DNA variation in the crested newt superspecies: limited cytoplasmic gene flow among species. *Evolution* 43: 88-104.

Walsh, P. S., D. A. Metzger and R. Higuchi, 1991. Chelex® 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* 10: 506-513.

Ward, R. H., B. L. Frazier, K. Dew-Jager and S. Pääbo, 1991. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Nat. Acad. Sci., USA* 88: 8720-8724.

Washbourn, C. K., 1967. Lake levels and Quaternary climates in the eastern Rift Valley of Kenya. *Nature* 216: 672-673.

Watterson, G. A., 1984. Allele frequencies after a bottleneck. *Theor. Pop. Biol.* 26: 387-407.

Wayne, R. K., N. Lehman, D. Girman, P. J. P. Gogan, D. A. Gilbert, K. Hansen, R. O. Peterson, U. S. Seal, A. Eisenhawer, L. D. Mech and R. J. Krumenaker, 1991. Conservation genetics of the endangered Isle Royale gray wolf. *Cons. Biol.* 5: 41-51.

Weber, A. W., 1987. Socioecologic factors in the conservation of Afromontane forest reserves. In *Primate Conservation in the Tropical Rain Forest*, edited by Mittermeier, R., p. 205-229. New York, Alan R. Liss.

Weir, B. S. and C. C. Cockerham, 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

Weiss, M. L., V. Wilson, C. Chan, T. Turner and A. J. Jeffreys, 1988. Application of DNA fingerprinting probes to old world monkeys. *Am. J. Primatol.* 16: 73-79.

Westerbergh, A. and A. Saura, 1994. Genetic differentiation in endemic Silene (Caryophyllaceae) on the Hawaiian Islands. *Am. J. Botany* 81: 1487-1493.

Whitmore, T. C., 1990. *An Introduction to Tropical Rain Forests*. New York, Oxford University Press.

Wickings, E. J. and A. F. Dixson, 1992. Application of DNA fingerprinting to familial studies of Gobonese primates. In *Paternity in Primates: Genetic Tests and Theories*, edited by Martin, R. D., A. F. Dixson and E. J. Wickings, p. 113-130. Basel, Karger.

Wildt, D. E., M. Bush, K. L. Goodrowe, C. Packer, A. E. Pusey, J. L. Brown, P. Joslin and S. J. O'Brien, 1987. Reproductive and genetic consequences of founding isolated lion populations. *Nature* 329: 328-331.

Wilkinson, L., M. Hill and E. Vang, 1992. *SYSTAT: Statistics*. Version 5.2. Evanston, IL, Systat, Inc.

Williams, M. A. J., 1973. Late Pleistocene tropical aridity synchronous in both hemispheres? *Nature* 253: 617-618.

Wilson, A. C., R. L. Cann, S. M. Carr, M. George, U. B. Gyllensten, K. M. Helm-Bychowski, R. G. Higuchi, S. R. Palumbi, E. M. Prager, R. D. Sage and M. Stoneking, 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* 26: 375-400.

Wolpoff, M., 1989. Multiregional evolution: the fossil alternative to Eden. In *The Human Revolution: Behavioral and Biological Perspectives on the Origin of Modern Humans*, edited by Mellars, P. and C. Stringer, p. 62-108. Princeton, Princeton University Press.

Woodruff, D. S., 1989. The problems of conserving genes and species. In *Conservation for the Twenty-first Century*, edited by Western, D. and M. C. Pearl, p. 76-88. New York, Oxford University Press.

Woodruff, D. S., 1992. Genetics and the conservation of animals in fragmented habitats. International Conference on Tropical Biodiversity: In Harmony with Nature, Kuala Lampur, Malaysia.

Wrangham, R. W., 1980. An ecological model of female-bonded primate groups.

Behaviour 75: 262-300.

Wrangham, R. W., 1986. Ecology and social relationships in two species of chimpanzee. In *Ecological Aspects of Social Evolution*, edited by Rubenstein, D. I. and R. W. Wrangham, p. 352-378. Princeton, Princeton University Press.

Wrangham, R. W., 1987. The significance of African apes for reconstructing human social evolution. In *The Evolution of Human Behavior: Primate Models*, edited by Kinzey, W. G., p. 51-71. New York, SUNY Press.

Wrangham, R. W., 1995. What accounts for the variety of chimpanzee traditions? Lecture given to the National Geographic Society.

Wrangham, R. W., C. A. Chapman and L. J. Chapman, 1994. Seed dispersal by forest chimpanzees in Uganda. *J. Tropical Ecol.* 10: 355-368.

Wrangham, R. W., C. A. Chapman, A. P. Clark-Arcadi and G. Isabirye-Basuta, 1996. The social ecology of Kanyawara chimpanzees: significance for the great apes. In *Great Ape Societies*, edited by McGrew, W. C. and T. Nishida, p. in press. Cambridge, Cambridge University Press.

Wrangham, R. W., A. P. Clark and G. Isabirye-Basuta, 1992. Female social relationships and social organization of Kibale Forest chimpanzees. In *Topics in Primatology: Human Origins*, edited by Nishida, T., W. C. McGrew, P. Marler, M. Pickford and F. B. M. de Waal, p. 81-98. Tokyo, University of Tokyo Press.

Wrangham, R. W., N. L. Conklin, C. A. Chapman and K. D. Hunt, 1991. The significance of fibrous foods for Kibale Forest chimpanzees. *Phil. Trans. Roy. Soc. Lond., Series B* 334: 171-178.

Wrangham, R. W., N. L. Conklin, G. Etot, J. Obua, K. D. Hunt, M. D. Hauser and A. P. Clark, 1993. The value of figs to chimpanzees. *Int. J. Primatol.* 14: 243-256.

Wrangham, R. W., F. B. M. de Waal and W. C. McGrew, 1994. The challenge of behavioral diversity. In *Chimpanzee Cultures*, edited by Wrangham, R. W., W. C. McGrew, F. B. M. de Waal and P. G. Heltne, p. 1-18. Cambridge, Harvard University Press.

Wrangham, R. W., W. C. McGrew, F. B. M. de Waal and P. G. Heltne, Eds., 1994. *Chimpanzee Cultures*. Cambridge, Harvard University Press.

Wright, S., 1931. Evolution in Mendelian populations. *Genetics* 16: 97-159.

Wright, S., 1943. Isolation by distance. *Genetics* 28: 114-138.

Wright, S., 1951. The genetical structure of populations. *Ann. Eugen.* 1: 323-334.

Wright, S., 1965. The interpretation of population structure by F-statistics with special regards to systems of mating. *Evolution* 19: 395-420.

Wright, S., 1969. *Evolution and the Genetics of Populations 2. The Theory of Gene Frequencies*. Chicago, University of Chicago Press.

Wroegmann, D., 1992. *Wild chimpanzees in Lopé, Gabon: Census Methods and Habitat Use*. Ph.D. thesis. Bremen University, Germany.

Wu, H. M., W. G. Holmes, S. R. Medina and G. P. Sackett, 1980. Kin preferences in infant *Macacca nemestrina*. *Nature* 285: 225-227.

Zink, R. M. and D. L. Dittmann, 1993. Gene flow, refugia and evolution of geographic variation in the song sparrow (*Melospiza melodia*). *Evolution* 47: 717-729.

Zischler, H., M. Höss, O. Handt and A. von Haeseler, 1995. Detecting dinosaur DNA. *Science* 268: 1192-1193.