INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI®

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Evolutionary Genetics of Chimpanzees (*Pan troglodytes*) in Nigeria and Cameroon

by

Mary Katherine Gonder

A dissertation submitted to the Graduate Faculty in Anthropology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2000

UMI Number: 9969690

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Anthropology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy

April 19, 2000
Date
Dr. John Oates, Chair of Examining Committee

April 19, 2000
Date
Dr. Louise Lennihan, Executive Officer

Dr. Sara Stinson

Dr. Ward Wheeler

Dr. Todd Disotell

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Evolutionary Genetics of Chimpanzees (*Pan troglodytes*) in Nigeria and Cameroon

by

Mary Katherine Gonder

Advisor: Professor John Oates

Chimpanzee populations have commonly been grouped into three subspecies: west African *Pan troglodytes verus*, central African *P. t. troglodytes* and east African *P. t. schweinfurthii*, with the western and central African subspecies said to be separated by the lower Niger River in Nigeria. The recent evolution and classification of *P. troglodytes* has been debated during the past decade as genetic evidence from wild chimpanzees has become available. Gonder *et al.* (1997) published mtDNA sequences from a limited sample of chimpanzees from Nigeria. These data suggested that a distinct chimpanzee lineage might be present in Nigeria and western Cameroon, and that the geographic barriers thought to delimit chimpanzees in western Africa might be different than proposed. This dissertation presents the most geographically comprehensive genetic database available for wild chimpanzees across Africa (n = 254), including 78 mtDNA d-loop sequences and 115 microsatellite genotypes of chimpanzees from many localities in Nigeria and Cameroon. Samples were collected on either side of putative biogeographic barriers throughout this region, including both sides of the Niger River.

These genetic data strongly suggest the need for recognizing two major lineages of chimpanzee, one in western Africa and one in western equatorial and eastern Africa. Chimpanzees in the Upper Guinea region, in Nigeria and in western Cameroon belong to a western African lineage. Chimpanzees in southern Cameroon, western equatorial Africa and eastern Africa belong to a central African lineage. The western African lineage might be further subdivided into two lineages, one represented by chimpanzees in the Upper Guinea region and one represented by chimpanzees in eastern Nigeria and western Cameroon. The results of this project do not support the view that chimpanzees in western equatorial and eastern Africa belong to two separate lineages.

The distributions of the western and central African lineages converge in Cameroon near the Sanaga River. The Sanaga River may have influenced chimpanzee distribution patterns for the last 100-200 thousand years, although a few migrants cross the Sanaga. The Niger River or the Dahomey Gap might separate chimpanzee populations in western Africa. These findings may partially be explained by the history of African forests during the Upper Pleistocene.

ACKNOWLEDGEMENTS

Many people have generously contributed to the completion of this project. First. I would like to thank my committee, John Oates, Todd Disotell, Sara Stinson and Ward Wheeler. I am grateful for their patience, understanding and encouragement while I was writing this document.

It has been my privilege to conduct this research under the direction of John Oates. I am grateful for the years I have spent as his student. John approached this project with a blend of skepticism about genetic research and a wide breadth of knowledge about African forests and the animals that live in them. His perspective has given this project a historical depth that it would have otherwise lacked. I thank John for introducing me to the forests of western Africa and for encouraging me to try to find chimpanzees in Nigeria.

I would also like to thank my unofficial advisor, Todd Disotell. Todd graciously and generously allowed me to hone my molecular skills and to pursue dissertation research in his laboratory at New York University. Todd's continuous enthusiasm and seemingly endless knowledge of all things molecular and analytical kept me motivated even during the darkest days of sample processing and analysis.

I am grateful to Sara Stinson for cheerfully and conscientiously attending committee meetings and reading drafts of my proposals and dissertation chapters. Thanks to Ward Wheeler for his valuable comments on theories and methods in molecular systematics.

I would also like to acknowledge the faculty and students of the New York Consortium in Evolutionary Primatology (NYCEP). I am pleased to have been included in such a dynamic academic community. Nearly every person involved with the NYCEP program has contributed directly or indirectly to the development and completion of this dissertation. Cliff Jolly and the students at the NYU laboratory also deserve recognition for their input on this project. I would especially like to thank Paul Telfer and Derek Wildman for maintaining my computer. I would also like to thank Don Melnick and the students at the Columbia University molecular laboratory for their assistance during the formative stages of this project. I especially thank Mike Forstner for teaching me how to sequence DNA.

Numerous organizations and people in Nigeria and Cameroon made the field phase of this project possible. I am especially grateful to Vassily and Claudia Barberopoulos for allowing me to stay in their home in Lagos, ride their horses and eat too much salad when I needed a break from the forest. Without their support, fieldwork in Nigeria would have been much more difficult. Thanks to Phil Hall of Pronatura for helping me obtain the necessary documents to enter Nigeria. Thanks to Andrew Dunn and John Barker of the World Wide Fund for Nature—U. K. for their assistance with logistical support at Gashaka Gumti and Cross River National Parks. Thanks to Dr. and Mrs. E. A. Agbelusi for helping me to obtain permits to conduct research in Ondo and Ekiti States. I would especially like to thank Peter Jenkins and Liza Gadsby of the *Pandrillus* Wildlife Rehabilitation Center for the hair samples of the Center's captive chimpanzees and for their logistical support in Calabar. I am grateful to Turner and Miriam Isoun of the Niger Delta Wetlands Foundation for their assistance with field work

in the Niger Delta. Thanks to the Nigerian Conservation Foundation for their assistance in Nigeria. Thanks to the numerous assistants that helped to locate chimpanzees, climb trees and collect samples. Finally, thanks to Salamu at Gashaka for helping me to get my first glimpse of wild chimpanzees.

In Cameroon, I would especially like to thank Bryan Curran and Dwight Lawson of the Wildlife Conservation Society. Without the logistical support and vehicle provided by WCS, research in Cameroon would have been very difficult. The staff at WCS--Cameroon was instrumental in helping me obtain the necessary CITES export permit documents. I would especially like to thank my project assistant, Monday Kassingang, for the wonderful job he did helping with all phases of research in Cameroon and for his friendship. Thanks to Steve Gartlan of the World Wide Fund for Nature--Cameroon for helping me to obtain the correct documents for a research permit. I am grateful to Glyn and Sylvia Davies for their help in coordinating sample collection on Mt. Cameroon. Thanks to the numerous field assistants in Cameroon who helped with locating chimpanzees and with sample collection. Without their help, I would still be lost in the forests of western Cameroon.

I am grateful to my family for their support. Thanks to my parents, Dan and Mary Gonder. Daddy's concern and encouragement kept me motivated during all phases of graduate student life. My mother set a high standard of excellence. I want to be just like her. Thanks to Ed Kujawski for his help in recompiling Phylip's Fast DNAml program. Thanks to Maggie, my dear friend and ambassador of good will, who watched me write every page of this document. Last, but far from least, I am grateful to my loving husband, Joe Kujawski, who knew that this project should be done.

This research was generously supported by the National Science Foundation (Dissertation Improvement Award, Graduate Research Fellowship and NYCEP Research Training Grant), the L.S.B. Leakey Foundation, Primate Conservation, Inc., and the Wenner-Gren Foundation. All samples were exported from Africa under CITES exportation permits (Nigeria: FEPA/LSN/68/T/39; Cameroon: 0172/PE/MINEF/DFAP/SL/SLP), and were imported under U.S. CITES (US810330) and USDA (97-418-2) importation permits.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

Chimpanzee Phylogenetics and Biogeography in Western Africa

Introduction

Chimpanzees (*Pan troglodytes*) are one of the most thoroughly studied and well-understood species in the world. Beginning with stories from a Carthaginian expedition led by Hanno in 500 B.C. reporting tailless human-like creatures in Africa (Ramin 1976) to Jane Goodall's groundbreaking work in chimpanzee socioecology at Gombe, Tanzania, during the 1960s and 1970s, interest in these close relatives to our own species remains unabated. Even though chimpanzees are well-understood compared to other species, the past decade has witnessed a proliferation of information suggesting that there is still much to learn about them; and that what has been accepted about their distribution patterns and recent evolution may, in fact, need to be revised.

The study of chimpanzee evolutionary biology has benefited enormously from studies using the field-based collection of hairs shed into the sleeping nests of wild chimpanzees. Using DNA isolated from hair, these studies have raised questions about intra- and inter-community relationships among chimpanzees, and have examined how forest dynamics, geography and demographic events have influenced their distribution and recent evolution. This study attempts to address some of these issues, particularly as they relate to the chimpanzee populations in present-day Nigeria and Cameroon.

Distribution and classification of chimpanzees

Three subspecies of chimpanzees (*Pan troglodytes*) are commonly recognized: west African *P. t. verus*, central African *P. t. troglodytes* and east African *P. t. schweinfurthii* (Figure 1.1). The widely accepted view of chimpanzee distribution patterns (Schwarz 1934; Hill 1967, 1969; Goodall 1986; Teleki 1989) suggests that:

1) *P. t. verus* ranges from Senegal to western Nigeria. *P. t. verus* is divided into two populations by the Dahomey Gap, a dry-forest zone covering present-day eastern Ghana, Togo and Benin.
2) The Niger River, in Nigeria, separates *P. t. verus* and *P. t. troglodytes.*
3) *P. t. troglodytes* ranges from eastern Nigeria to the Ubangi River in the Democratic Republic of Congo (the former Zaire), and as far south as the Congo River.
4) The Ubangi River separates *P. t. troglodytes* and *P. t. schweinfurthii.*
5) *P. t. schweinfurthii* ranges from the Ubangi River and as far east as the western Rift Valley.

This picture of chimpanzee distribution patterns has been reinforced by studies conducted from the 1960s into the early 1990s. Craniometric studies suggested that three kinds of chimpanzee are present in museum collections, and that they follow Schwarz's (1934) geographic patterning (e.g., Shea and Coolidge 1988; Groves, Westwood, and Shea 1993). Studies of pelage characteristics and analyses of protein variation in the chimpanzees housed at Holloman Air Force Base reported similar findings (Goodman *et al.* 1967; Hill 1967; Goodman and Tashian 1969; Hill 1969); although there have been notable exceptions (Reynolds and Luscombe 1970). Morin *et al.* (1994) were the first to use DNA derived from field-collected shed hair. While analyses of their pan-African

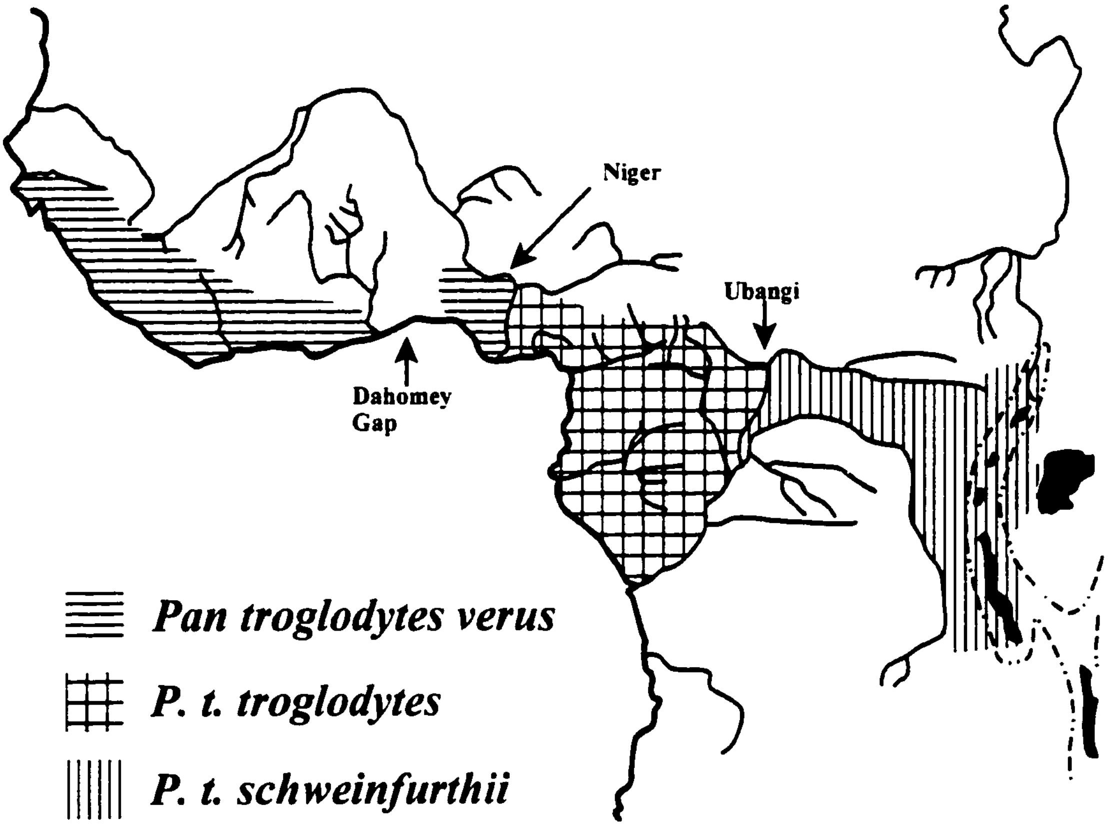


Figure 1.1. Chimpanzee distribution after Schwarz (1934) and Hill (1967, 1969). *P. t. koolakamba* is excluded from this distribution map due to a lack of evidence to support its existence.

sample suggested that there are three major chimpanzee lineages, they were the first to report that recent chimpanzee evolution may differ from the widely accepted scenario.

All of these studies, however, suffer from a common problem: the geographic origins of their samples were largely unknown, inhibiting analyses on a fine geographic scale. The craniometric analyses relied upon the accurate recording of provenance information in field-notes taken during expeditions. These notes can be inaccurate and misleading. Similarly, records kept for captive animals are notoriously unreliable. Data sets included in all of these studies were also geographically incomplete, lacking samples from many areas and not including samples collected near proposed geographic boundaries. Morin *et al.*'s (1994) samples of known provenance, for example, largely came from Taï in Ivory Coast and Gombe in Tanzania; the origins of samples from chimpanzees from elsewhere in Africa were usually known only in broad terms, if at all.

A brief history of chimpanzee nomenclature

Over the last 150 years, chimpanzee variation and distribution patterns have received considerable attention in the literature. From the 1850's until the 1930's, several researchers published their own distribution and nomenclature systems. Their conclusions were based largely on skins and skeletal remains collected during hunting and specimen-collecting expeditions led by naturalists and explorers. During that time, nearly every new specimen added a new chimpanzee taxon, especially for specimens collected in western equatorial Africa.

Du Chaillu (1861), Gray (1860, 1861, 1862, 1867 and 1870), Reichenbach (1862) and Gratilet and Alix (1866) were the first to document intraspecific differences among chimpanzees (Hill 1967, 1969). Each author proposed different generic and specific names, based on the examination of a few specimens collected during field expeditions. It was not until the early 20th century that systematic attempts to collate this piecemeal information were published. Table 1.1 summarizes major nomenclature schemes of chimpanzees from 1904 to 1969. Figure 1.2 depicts the geographical distribution of the named taxa.

Matschie (1904) investigated multiple samples from different putative chimpanzee populations (Hill 1967, 1969). Using the generic name *Simia*, he recognized seven distinct species of chimpanzee. In his distribution, six of these species ranged in Upper Guinea and western equatorial Africa; and one ranged eastern Africa. In two subsequent papers, Matschie (1914, 1919) recognized no less than fourteen chimpanzee species, but these publications were largely ignored in favor of his earlier work (Hill 1969).

Rothschild (1905, 1906) published his nomenclature and distribution scheme based on examinations of over 100 museum specimens and a few captive specimens living in Europe. Like Matschie, Rothschild based his work on examining coat coloration and skin pigment variation. He also added several crania to his sample, and from them deduced overall differences between groups. Using the generic name *Simia*, Rothschild recognized five species and seven subspecies of chimpanzee. The type specimens for each of the species he recognized were derived from his western equatorial Africa sample. Subspecies of *Simia pygmaeus* and *S. satyrus* represented chimpanzees in western and eastern Africa, respectively.

Table 1.1. Major chimpanzee nomenclatures

	Nomenclature*	Distribution	Author
Matschie 1904			
1	*Simia satyra*	Gaboon (type locality)	Linnaeus, 1758
2	*S. calvus*	Mt. Cameroon	Du Chaillu, 1860
3	*S. vellerosus*	Interior of Gaboon	Gray, 1862
4	*S. schweinfurthii*	East Africa (Niam-niam country)	Giglioli, 1872
5	*S. fuscus*	Liberia and Togoland	Meyer, 1894
6	*S. leucoprymnus*	Guinea Coast	Lesson, 1831
7	*S. chimpanse*	Gambia and Senegambia	Mayer, 1856
Rothschild 1905, 1906			
8	*S. vellerosus*	Cameroons	Gray, 1862
9	*S. v. fuliginosus*	Congo	Schaufuss, 1870
10	*S. satyrus*	Cameroons and Gaboon	Linnaeus, 1758
11	*S. s. marungensis*	Central Congo	Noack, 1897
12	*S. s. schweinfurthii*	Sudan and Uganda	Giglioli, 1872
13	*S. koolokamba*	Cameroons and Gaboon	Du Chaillu, 1860
14	*S. aubryi*	Cameroons and Gaboon	Gratiolet and Alix, 1866
15	*S. pygmaeus*	Congo	Schreber, 1796
16	*S. p. fuscus*	Gold Coast	Meyer, 1894
17	*S. p. leucoprymnus*	Sierra Leone and S. Liberia	Lesson, 1831
18	*S. p. chimpanse*	Gambia	Matschie, 1904
19	*S. p. raripilosus*	French Congo	Rothschild, 1905

Table 1.1. Major chimpanzee nomenclatures (continued)

	Nomenclature*	Distribution	Author
Elliot 1913			
20	*Pan calvus*	Cameroon and Gaboon	Du Chaillu, 1860
21	*P. fuliginosus*	Gaboon and French Congo	Schaufuss, 1870
22	*P. satyrus*	Gaboon	Linnaeus, 1758
23	*P. koolokamba*	Cameroon and Gaboon	Du Chaillu, 1860
24	*P. leucoprymnus*	?Sierra Leone and Liberia	Lesson, 1831
25	*P. chimpanse*	Gambia	Mayer, 1856
26	*P. schweinfurthii*	Eastern Africa (Niam-niam country, Albert Nyanza west to Tanganyika and into the Congo Forest). Differences in distribution of *P. schweinfurthii* and *P. s. marungensis* were not given	Giglioli, 1872
27	*P. s. marungensis*		Noack, 1887
28	*P. aubryi*	Cameroon and Gaboon	Gratiolet and Alix, 1866
29	*P. vellerosus*	Cameroon	Gray, 1862
30	*P. fuscus*	Gold Coast	Meyer, 1894, 1895
Schwarz 1934			
31	*Pan troglodytes verus*	Upper Guinea	Schwarz, 1934
32	*P. t. satyrus*	Lower Guinea	Blumenbach, 1779
33	*P. t. schweinfurthii*	East Africa	Giglioli, 1872
34	*P. t. paniscus*	Left Bank of the Congo River	Schwarz, 1929
Hill 1967, 1969			
35	*Pan troglodytes verus*	Upper Guinea Forest Block	Schwarz, 1934
36	*P. t. troglodytes*	Lower Guinea Forest Block	Blumenbach, 1779
37	*P. t. koolokamba*	Lower Guinea Forest Block	Du Chaillu, 1860
38	*P. t. schweinfurthii*	East Africa	Giglioli, 1872

*Criteria for naming new species given in Wilson and Reeder (1993).

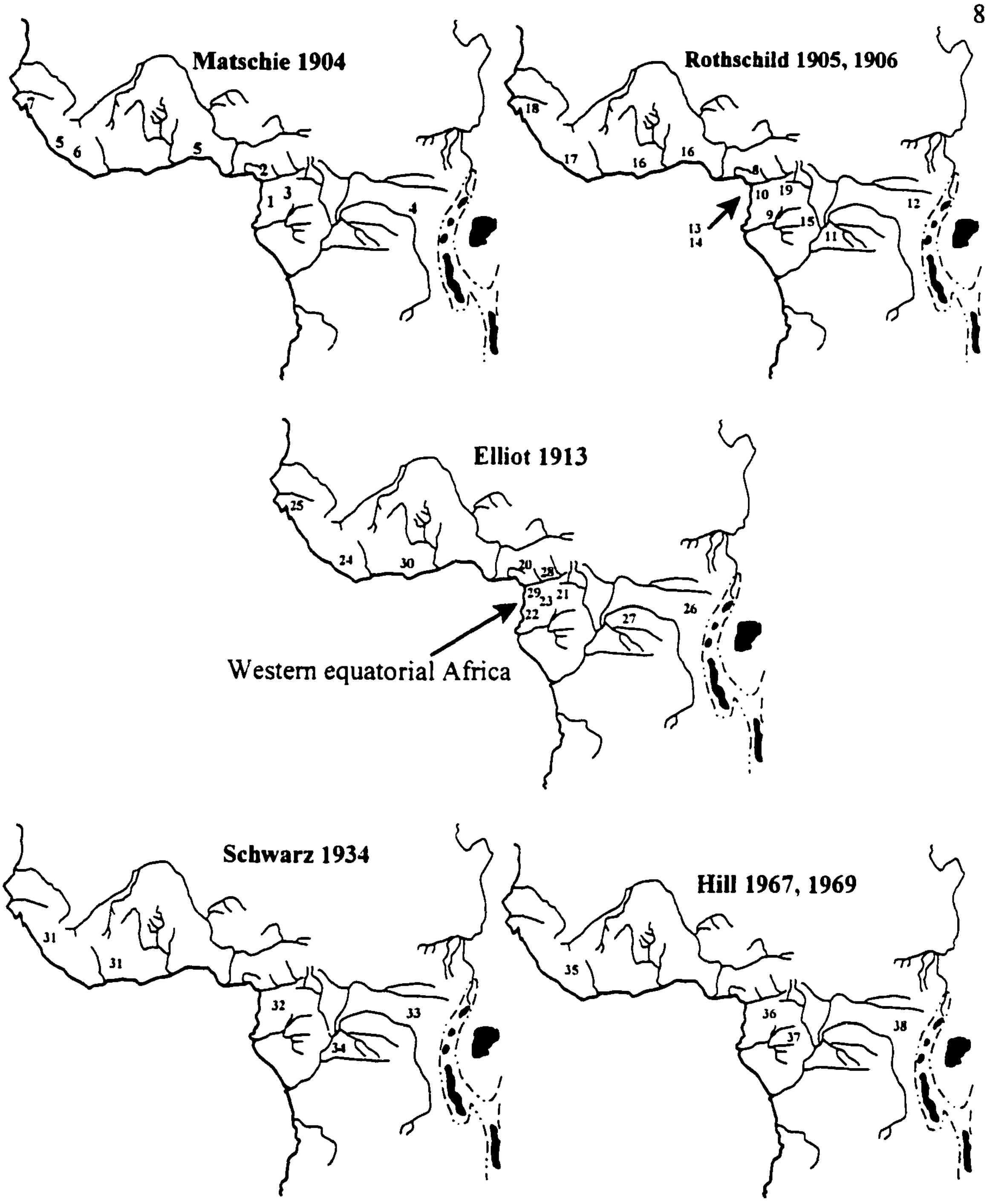


Figure 1.2. Geographic distribution of type specimens used to describe chimpanzees. Numbers correspond to taxon names given in Table 1.1. Note that most named taxa are located in western equatorial Africa.

Elliot (1913) argued that the specimens found in European museums lacked intermediate forms between recognized taxa. In addition, he strongly argued that species are stable units, and rejected the idea that the numerous subspecies recognized by other taxonomists could exist in the natural world. Consequently, he presented a classification scheme that included ten species and one subspecies, and used the generic name *Pan*. Elliot's classification was a hybrid of Matschie's (1904) and Rothschild's (1905, 1906) classifications. He elevated all the specimens to full-species status previously named by Matschie and Rothschild. He retained only a single subspecies from Rothschild's (1905, 1906) classification, *Pan schweinfurthii marungensis*.

Convinced that the number of chimpanzee taxa described far exceeded the number of distinct forms that actually existed, Schwarz (1934) presented a new classification that collapsed most of the previously described taxa. Acknowledging that a considerable amount of variation is present among chimpanzees, Schwarz recognized only one species, *Pan troglodytes*, and four subspecies. *P. t. verus* included all previously designated western African taxa (taxa 5-7, 16-18, 24, 25 and 30 on Figure 1.2); *P. t. troglodytes* included most previously designated western equatorial African taxa (taxa 1-3, 8-10, 13-15, 20-23, 28 and 29 on Figure 1.2); and *P. t. schweinfurthii* included most previously recognized eastern African taxa (taxa 4, 12 and 26 on Figure 1.2). The fourth subspecies, *P. t. paniscus*, represented specimens that had been recently obtained from chimpanzee populations south of the Congo River. Schwarz (1934) also proposed that large rivers (the Niger, Ubangi and Congo Rivers) separate these subspecies.

In the previous year, Coolidge (1933), presented his argument that there was sufficient variation from other chimpanzees present in the *P. t. paniscus* sample to warrant full-species status for those specimens. Coolidge (1933) named them *P. paniscus*. Subsequent to those publications, 36 years passed with the only real debate surrounding the taxonomic status of *Pan paniscus*.

Hill (1967, 1969) revised Schwarz's classification to include two chimpanzee species: *Pan troglodytes* and *Pan paniscus*. In the *P. troglodytes* sample he recognized: west African *P. t. verus*, central African *P. t. troglodytes*, east African *P. t. schweinfurthii*, and introduced a new subspecies, *P. t. koolakamba* (taxa 2, 12 and 23 on Figure 1.2). Hill believed that a *P. t. koolakamba* specimen was present in the chimpanzee colony at Holloman Air Force base. He claimed that the particular animal in question (picture in Hill 1969, 42) displayed the phenotypic characteristics and savage disposition of a gorilla-chimpanzee hybrid that had been described by local informants around the head waters of the Ogowe River near the Ashankolo Mountains in western equatorial Africa. Recent studies have failed to find support for this subspecies (Shea and Coolidge 1988; Groves, Westwood, and Shea 1993).

This very brief review of chimpanzee nomenclature suggests several interesting properties of this species. Although it was fashionable during the 19th and early 20th centuries to split populations into as many groups as possible, these classifications suggest that great genetic, morphological, behavioral and ecological diversity is present among wild chimpanzees, a fact that several recent studies have confirmed (behavior and ecology: McGrew *et al.* 1997; Mitani, Hunley, and Murdoch 1999; Whiten *et al.* 1999). These classification schemes imply also that most chimpanzee genetic and morphological diversity is centered on western equatorial Africa. The majority of all specimens used to differentiate chimpanzee taxa were found in this region (Table 1.1 and Figure 1.2). In fact, Elliot observed:

> If our determinations are correct, we have the singular fact that nearly all the recognized forms of chimpanzees...are crowded together on a small portion of West Africa [western equatorial Africa], leaving us to wonder how so many distinct forms, if they are such, could exist in so restricted a territory and preserve their specific characters intact (Elliot 1913, 233).

Finally, this review suggests that chimpanzee distribution patterns, and the geographic features thought to limit the range of chimpanzees, were proposed with little to no evidence. Even Hill (1967, 1969) noted that Schwarz's nomenclature probably was oversimplified. In fact, Schwarz did not provide any evidence supporting the Niger or Ubangi Rivers as boundaries to chimpanzee distribution.

African forest history and biogeography

What may have fostered such a high level of genetic and phenotypic diversity among chimpanzees in western Africa? One reason may be the complex history of the forests in this region. Despite the inevitably different histories and dispersal abilities of separate lineages, African rainforest primates tend to cluster into discrete communities which are believed to reflect a shared historical relationship with the forests they inhabit (Haffer 1969; Endler 1982; Haffer 1982; Mayr and O'Hara 1986; Kingdon 1989). Rivers, mountain ranges and possibly interspecific competition further complicate the evolutionary history of particular species (Booth 1958a, 1958b; Schiøtz 1967; Moreau

1969; Robbins 1978; Grubb 1982; Oates 1988; Grubb 1990). Each of these factors may have acted separately or in tandem to produce the patterns of chimpanzee morphological and genetic diversity observed in today in western Africa. I will discuss these factors in the following sections.

African forest history

Earth's climate has varied considerably over the last 2.8 million years. This variation is attributed to cyclic fluctuations in Earth's orbital parameters (Imbrie and Imbrie 1980; Cox and Moore 1993; Rial 1999). Fluctuations in Earth's orbital parameters produce cyclic expansions and contractions of the polar ice caps and ice sheets (Imbrie 1985; Hamilton 1988; Bonnefille, Roeland, and Guiot 1990; Hamilton 1992; Imbrie *et al.* 1992; Livingstone 1993; Maley 1996). Figure 1.3 illustrates the variation in Arctic ice cap volume during the Plio-Pleistocene. These data suggest that from roughly 2.8 to 1 million years ago (mya) pronounced cyclic changes occurred in ice-cap volume with a dominant 41 thousand-year (ky) periodicity. Beginning roughly 800 thousand years ago (kya), the amplitude of these cycles increased, developing a dominant 100ky periodicity (Imbrie and Imbrie 1980; Imbrie *et al.* 1992; deMenocal 1995; Maley 1996).

Volume changes in the Antarctic ice sheets parallel the changes of the Arctic ice cap (Lowell *et al.* 1995; Maley 1996; Mayewski *et al.* 1996; Conway *et al.* 1999; Fischer *et al.* 1999). It is not entirely clear what synchronizes these cycles between the poles. Atmospheric greenhouse gases may interact with ocean currents, which in turn

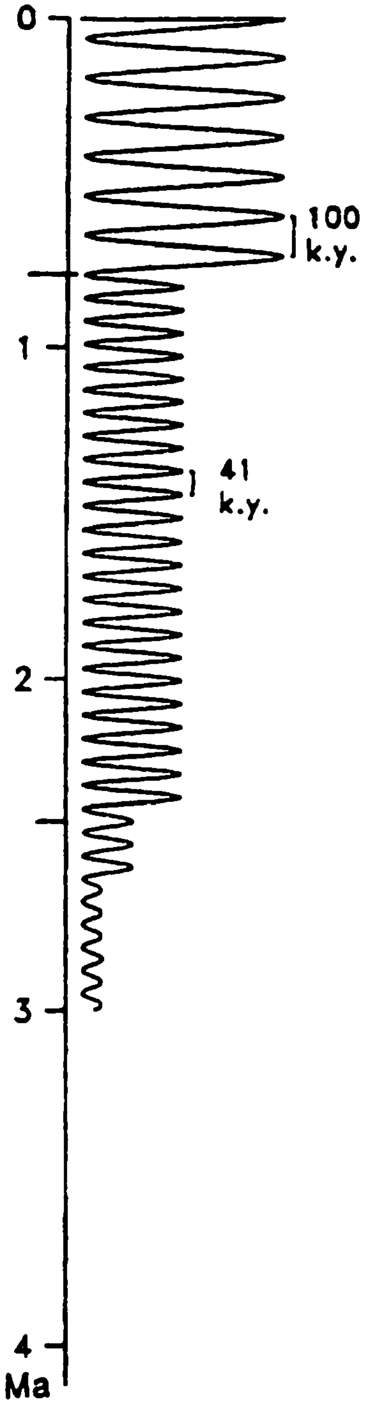


Figure 1.3. Schematic representation of the major cycle variations in the Arctic ice cap volume over the last three million years (redrawn from Maley 1996). Note the onset of cycling glacial conditions beginning 2.8 million years ago and the increase in amplitude and periodicity beginning 800 thousand years ago. The Arctic ice volume cycles reflect global patterns inferred from several independent paleoclimatic data sets.

synchronize polar ice volume changes (Livingstone 1993; Maley 1996; Fischer *et al.* 1999; Fluckiger *et al.* 1999). The history and extent of ice cap and ice sheet expansion has varied considerably between locations (Imbrie *et al.* 1992; Lowell *et al.* 1995; Lang *et al.* 1999). However, at times the polar ice caps and high-latitude ice sheets have extended several thousand kilometers beyond their present limits (Livingstone 1975; Imbrie 1985; Imbrie *et al.* 1992; Livingstone 1993).

Several independent data sets, representing a broad geographic scale, reveal that global climate was affected by these high-latitude ice volume cycles. Levels of atmospheric gas isotopes, atmospheric dust, and pH variation in fossil ocean microorganisms derived from analyses of polar ice cores, montane ice cores and deep-sea cores indicate that Earth's temperature and climate closely tracked these cyclic changes in polar ice volume (Livingstone 1975; Hamilton 1992; Livingstone 1993; deMenocal 1995; Maley 1996; Palmer, Pearson, and Cobb 1998; Thompson *et al.* 1998). This cyclic pattern also is well documented by other lines of data, including global lake and sea level changes, and shifts in sand dune geomorphology (Bonnefille, Roeland, and Guiot 1990; Maley 1991; Livingstone 1993; Gallup, Edwards, and Johnson 1994; deMenocal 1995; Maley 1996; Nichol 1999).

Overall, these data indicate that periods of increased glaciation closely correspond to lower average global temperature (Bonnefille, Roeland, and Guiot 1990; Imbrie *et al.* 1992; Livingstone 1993; deMenocal 1995; Maley 1996). Water and atmospheric moisture were trapped in the expanding ice sheets, making these periods cooler and subject to aridity. As the ice sheets receded, water was released into the atmosphere, rendering the warmer inter-glacial periods more humid (Imbrie 1985; Hamilton 1992;

Livingstone 1993; deMenocal 1995; Maley 1996). Transitions from cooler conditions to warmer conditions were abrupt, and may have occurred over the span of only a few centuries (Eglinton *et al.* 1992; Oppo, McManus, and Cullen 1998; Lang *et al.* 1999). However, the relationship between glacial phases and atmospheric humidity is complex. Prevailing local climatic conditions also influenced the degree of moisture in the atmosphere and the subsequent climatic response during glacial and inter-glacial phases (Imbrie 1985; Maley 1996; Fischer *et al.* 1999).

Substantial evidence suggests that local climate in Africa closely tracked shifts in global temperature (Livingstone 1975; Hamilton 1992; Livingstone 1993; deMenocal 1995; Maley 1996). This evidence suggests that, during periods of maximum glaciation, local climate in Africa was cooler and drier than at present. These shifts towards a cooler, drier African climate are primarily attributed to the upwelling of cold deep-sea water off the coast of northwestern and southwestern Africa that was pushed towards the equator during periods of increased glaciation. The lower sea surface temperature in turn reduced evaporation rates, leading to weakened southwesterly monsoons across western and western equatorial Africa (Hamilton 1992; Maley 1996). This chain of events may have reduced the temperature in equatorial Africa by as much as 6 to 10°C, and decreased rainfall by up to 30% (Bonnefille, Roeland, and Guiot 1990; Livingstone 1993; Maley 1996).

Over the last 2.8 million years there have been roughly 21 major fluctuations in global climatic temperature (Hamilton 1992). Although the data are the most complete for the last 75ky, it reasonable to assume that earlier climatic shifts had similar effects on African climate (Maley 1996). Over the last 250ky, three significant glacial episodes have occurred, with less extreme cooling trends punctuating interglacial periods. Approximately 250kya an arid period began in Africa that lasted roughly 50ky, followed by a 10ky warm, wet interglacial period. Another cool, arid advance occurred between 190-155kya followed again by a relatively warm interglacial period from about 128kya to 75kya. This interglacial period was interrupted by two short less intense cooling phases. After 75kya, another cooling phase began. This period came to an abrupt end about 10kya, and was followed by a less extreme cooling period during the mid-Holocene (Kingdon 1989; Hamilton 1992; deMenocal 1995; Maley 1996).

Several paleoclimatic data sets reflect trends towards increased aridity across sub-Saharan Africa during periods of maximum glaciation. Atmospheric dust deposits and isotopic data from deep-sea cores taken from many localities along the African coast suggest a much more open and arid climate during periods of maximum glaciation (Maley 1991; deMenocal 1995; Maley 1996). Lake-levels in sub-Saharan Africa were considerably depressed (Talbot and Delibrias 1977; Street and Grove 1979; Talbot *et al.* 1984; Maley 1991, 1996); and sand dunes occurred in areas presently too humid to permit their formation (Livingstone 1993; Nichol 1999). Montane glaciers also expanded below their present levels. During the last glacial advance, for example, montane glaciers in eastern Africa descended over a thousand meters below their present limits (Hamilton 1987; Osmaston 1989).

The vegetation of Africa changed concomitantly with global climatic shifts. Deep-sea cores and lake sediments indicate that during periods of maximum glaciation, more open habitats were in place, along with an abundance of herbaceous species adapted to an arid climate (Hamilton 1992; Livingstone 1993; Maley 1996). Atmospheric dust levels across Africa also increased during these periods (deMenocal 1995). Conversely, during inter-glacial periods these records document an increase in pollen from species adapted to a more humid forested environment and lower levels of atmospheric dust, implying a re-expansion of forests into formerly-arid habitats (Hamilton 1992; Livingstone 1993; Maley 1996). At the end of the last glaciation, the tropical forest reached its maximum expanse during the mid-Holocene. After 7kya, a less extreme cooling period combined with the increasing influence of widespread agriculture (beginning about 3kya) has reduced the overall extent of the forested area established across Africa during the mid-Holocene (Livingstone 1975; Kingdon 1989; Maley 1991; Hamilton 1992; deMenocal 1995; Maley 1996).

Forest history in western Africa

Like other regions of sub-Saharan Africa, the vegetation of western Africa shifted towards more open and arid habitats during periods of maximum glaciation. These periods were followed by a re-expansion of forest vegetation after the amelioration of climate. Deep-sea core data parallel lake-level and sediment data from Lake Bosumtwi in Ghana (Talbot and Delibrias 1977; Talbot *et al.* 1984). These data document shifts from arid-adapted herbaceous species to moisture-dependent forest species and changes in atmospheric dust levels that closely tracked shifts in global temperature and humidity (Livingstone 1993; deMenocal 1995; Maley 1996). The level of Lake Bosumtwi was considerably depressed during the last glacial maximum (Talbot and Delibrias 1977; Talbot *et al.* 1984); and ancient sand dunes are present in areas of western Africa currently supporting tropical forests (Nichol 1999).

The palynological data from Lake Bosumtwi cover the period from 28ky to 8.5ky (Talbot and Delibrias 1977; Talbot *et al.* 1984). These data document a transition from montane forest species adapted to a cool, wet environment to a marked increase in pollen species associated with medium altitude montane prairie from 28kya to 19kya. Less than 5% of the pollen record consists of forest tree species at the end of that period. After 15kya, the pollen record documents a transition from medium altitude prairie pollen to pollens representing forest-adapted species. After 8.5kya, the pollen record contains a predominance of forest species, implying a large re-expansion of the forest (Talbot and Delibrias 1977; Talbot *et al.* 1984; Maley 1996).

A marked contrast is evident in the palynological records from Lake Barombi Mbo in western Cameroon (Maley *et al.* 1990; Maley 1996). Between 20kya and 10kya the palynological record documents an increase in montane forest and prairie vegetation in the area. However, the occurrence of these species was not as pronounced as in the Lake Bosumtwi pollen records. Forest tree species still accounted for over 50% of the species represented. After 10kya, the palynological record of Lake Barombi-Mbo documents a predominance of forest species. These data demonstrate that forest species persisted in southeastern Nigeria and western Cameroon during the last glacial maximum, but further to the west near Lake Bosumtwi, forest cover was more greatly reduced (Maley *et al.* 1990; Maley 1991, 1996).

It is often stated that the last glacial advance was the most extreme (Imbrie 1985; Hamilton 1992; Livingstone 1993; deMenocal 1995; Maley 1996). However, western Africa may be different from other regions in that respect. Nichol (1999) has analyzed ancient sand-dune formations in central Nigeria (approximately 7°N) and compared them to ancient sand dunes from northern Nigeria dated at 40kya. Her comparisons of the wear and relief patterns of these ancient sand dunes reveal that that those in central Nigeria are older than those further to the north. Consequently, she has proposed that the ancient sand dunes in central Nigeria are the product of an earlier glacial advance during the mid-Pleistocene (either 250, 190-155 or 120-90 kya). Given the southerly extent of those ancient sand dunes, Nichol hypothesizes that during at least one mid-Pleistocene glacial episode, most of the forest in western Africa, and possibly in western equatorial Africa, was completely destroyed.

African biogeography

Several authors have noted that certain areas of tropical Africa contain many more species and more endemic taxa than other areas (Haffer 1969; Endler 1982; Haffer 1982; Mayr and O'Hara 1986). Consequently, these investigators have suggested that forest animals must have persisted in those areas during the cool, arid phases of the Pleistocene. This reasoning forms the basis of the well-known Pleistocene refuge hypothesis (Endler 1982; Haffer 1982; Mayr and O'Hara 1986; Hamilton 1988, 1992). This hypothesis suggests that during periods of maximum glaciation, tropical rainforests persisted only in limited areas with suitable prevailing local climatic conditions. It further suggests that, as their former ranges contracted, forest animals were isolated in these refuges. Following the amelioration of climate, some taxa reclaimed their former ranges as the forest was re-established.

Forest lineages that persisted may have differentiated due to the isolation of local populations in refuges (vicariance) and/or due to events accompanying the dispersal of these species during post-glacial forest expansion (Grubb 1982; Hamilton 1988; Kingdon 1989; Grubb 1990). Both forces, vicariance and dispersal, may have influenced the current distributions of the African tropical forest fauna. However, the current distribution patterns of the African forest fauna are too complex to be explained entirely by isolation in, or dispersal out of, refuges produced by the last glacial advance. Instead. refuges probably have had a significant influence on forest animal species throughout the series of Pleistocene glacial cycles (Hamilton 1988, 1992; Livingstone 1993: Maley 1996).

The locations of Pleistocene refuges are ambiguous; they have been inferred retrospectively, based on an area's level of species richness and species endemism. Areas of high species diversity and those containing more endemic species are often proposed as refuges (Grubb 1982: Chapman 1983: Mayr and O'Hara 1986: Hamilton 1988: Kingdon 1989; Grubb 1990: Colyn, Gautier-Hion, and Verheyen 1991: Hamilton 1992: Cox and Moore 1993). Hamilton (1988) suggested five large core refuge areas were present across Africa in: Sierra Leone/Liberia , eastern Ivory Coast/western Ghana, Cameroon/Gabon, the eastern Congo Basin and eastern Tanzania (Figure 1.4a).

While these areas probably did harbor refugia, the history and locations of the refuge areas appears to have been very complex (Figure 1.4b). These areas probably were not simply large tracts of forest that persisted despite drier global climatic conditions. In addition, smaller refuges may have persisted outside each core area, making the history of forest animal lineages more difficult to reconstruct (reviewed in Maley 1991, 1996). For example, in a reanalysis of the distributions of several primates and small mammals in the Congo Basin, Colyn, Gautier-Hion and Verheyen (1991)

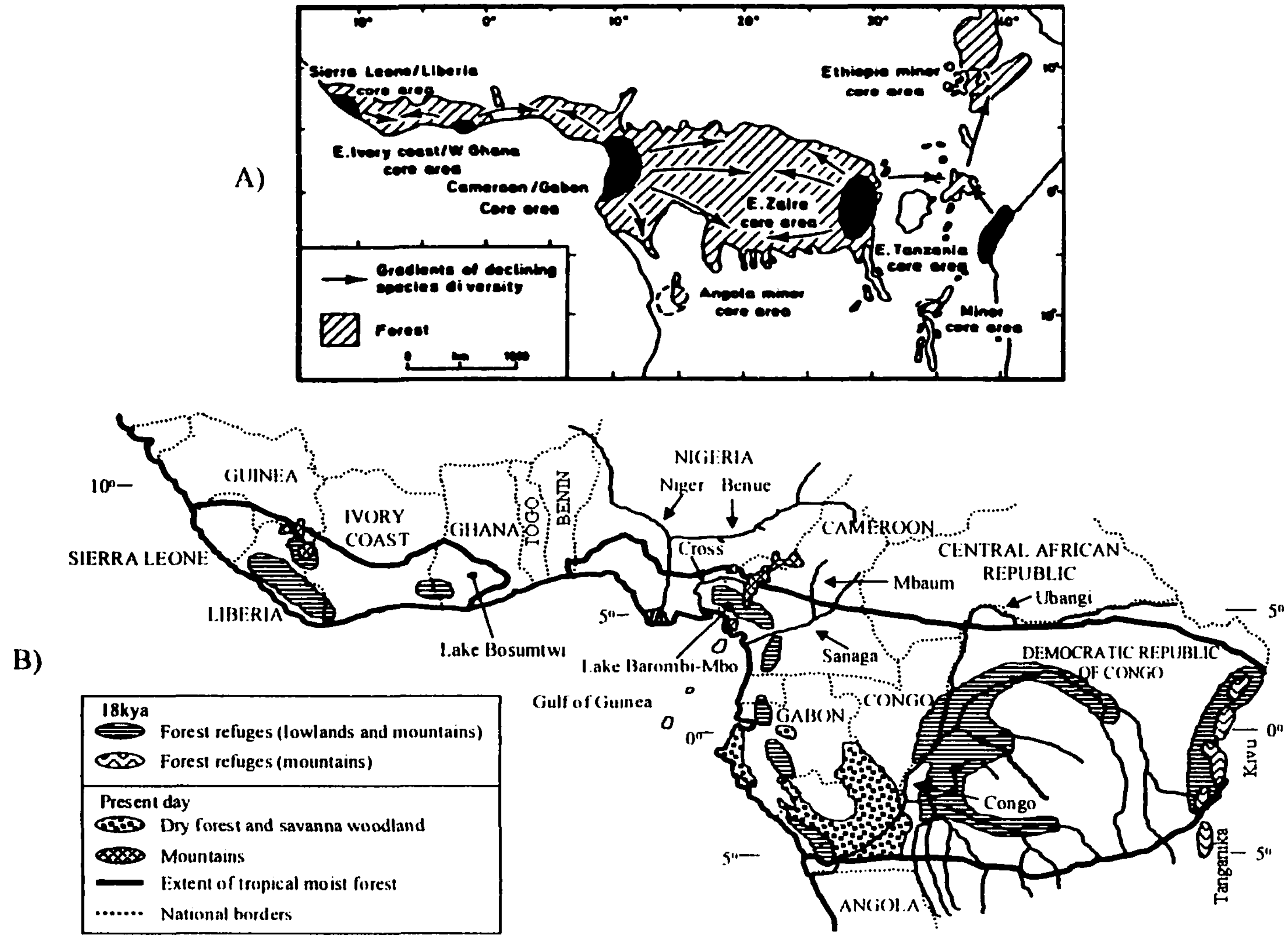


Figure 1.4. Pleistocene forest refugia. The top panel illustrates traditional locations of large core refuge areas inferred from centers of high species' endemism (redrawn from Hamilton 1988). The bottom panel shows a consensus of the probable locations and configuration of forest refugia (redrawn from Maley 1996).

proposed a redefinition of the Congo Basin core area. Their evidence suggests that a series of small, interconnected fluvial refuges that followed the course of the tributaries of the Congo River may have persisted, rather than a single large core area. Smaller refuges probably persisted outside this area, particularly along the eastern African coast and on large mountains, such as, Mt. Kilamanjaro and Mt. Elgon (Hamilton 1987, 1988; Kingdon 1989; Hamilton 1992; Livingstone 1993).

Biogeography in western Africa

Refuges. Analyses of the animal distribution patterns of the western and western equatorial African tropical forest taxa support the idea that forest refuges probably persisted in Liberia/Sierra Leone, western Ghana and Gabon (Grubb 1982; Mayr and O'Hara 1986; Hamilton 1988; Kingdon 1989; Grubb 1990; Hamilton 1992). As in eastern Africa, the history of the refuges and of the lineages evolving within them appears to have been more complex than can be explained entirely by the persistence of a few large core forested refuges. For instance, several authors have proposed that isolated refuges lay outside the larger core areas (reviewed in Maley 1991, 1996).

The evolutionary history of the forest and of forest animals must have been particularly complex in the vicinity of Nigeria and Cameroon. The region is highly speciose and geographically complex. Among the primates, for example, this area is often proposed as an evolutionary hot spot, with high species diversity and high species endemism (Kingdon 1989; Sayer, Harcourt, and Collins 1992; Hacker, Cowlishaw, and Williams 1998). Consequently, some investigators have proposed that several disjunct refuges may have persisted in the area (Figure 1.4b). The Niger delta may have harbored a small isolated refuge (Moreau 1969; Hamilton 1988; Kingdon 1989; Hamilton 1992), although it is not illustrated by Maley (1996). Several authors have analyzed the complex distributions of forest taxa from eastern Nigeria and into Gabon (Grubb 1982; Hamilton 1988; Oates 1988; Kingdon 1989; Grubb 1990; Maley 1996). Overall, the distribution patterns of these forest taxa imply that two refuges may have existed in this region. The northern refuge may have covered the area east of the Cross River to Mount Cameroon and on to the Sanaga River; while the more southerly refuge persisted south of the Sanaga River into Gabon (see Figure 1.4b).

Dispersal barriers. The distribution patterns of forest species observed across tropical Africa are more complex than can be fully explained by isolation in and dispersal out of Pleistocene refuges, especially in western Africa (Grubb 1982; Chapman 1983; Oates 1988; Grubb 1990). Both the vicariant and dispersal models require the presence of barriers that influence species dispersal and gene flow patterns (Oates 1988; Cox and Moore 1993). Consequently, several investigators have proposed that geographic barriers, such as rivers and mountain ranges, may have further influenced the distribution patterns of the rainforest fauna.

Across western Africa, several such barriers have been proposed (Figure 1.5). Featuring prominently in the literature are vegetational discontinuities in the forest cover, most notably the Dahomey Gap, and to a lesser extent the Baoulé-V. Several rivers also have received a great deal of attention. To the west, the Cavally, Sassandra, Bandama and Comoé Rivers have been proposed to limit the distributions of several taxa.

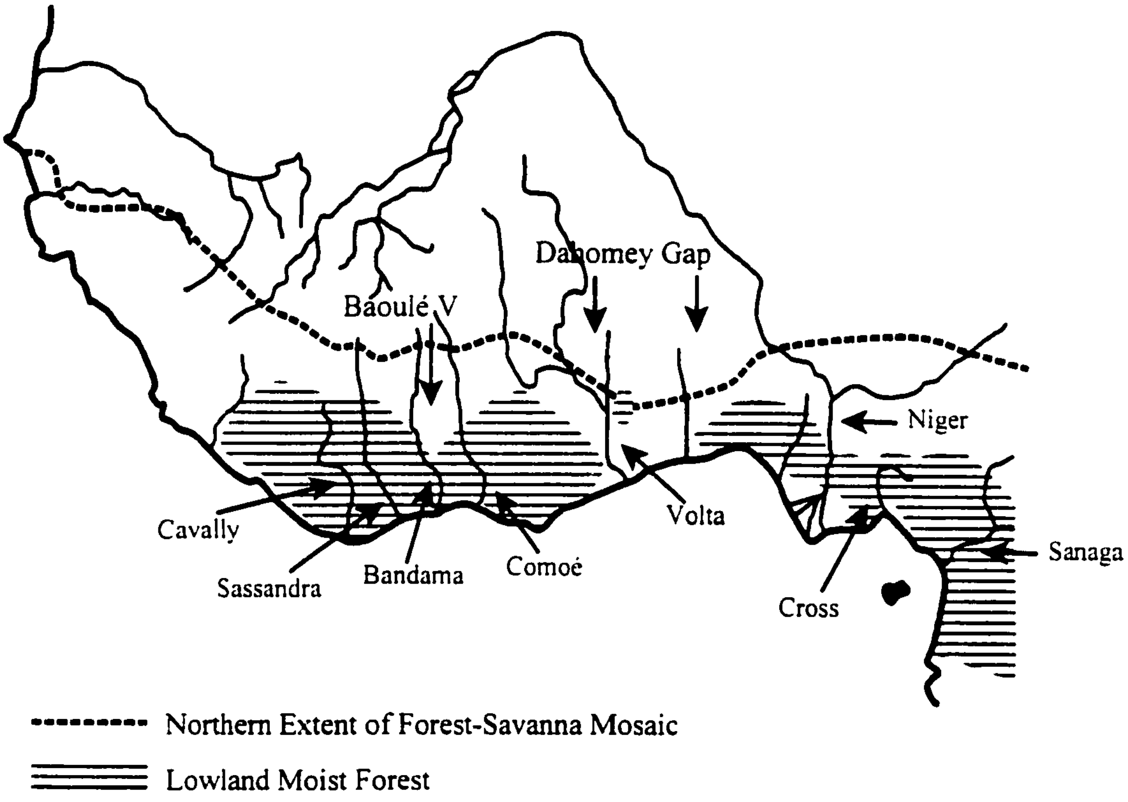


Figure 1.5. **Proposed dispersal barriers in western and western equatorial Africa.**

Further to the east the Volta, Niger, Cross and Sanaga Rivers may have been important in the history of many lineages. In the Nigeria-Cameroon region, the Dahomey Gap, the Niger, Cross and Sanaga Rivers, as well as the Cameroon Highlands have been proposed to have influenced the distribution of numerous organisms (Booth 1958a, 1958b; Schiøtz 1967; Moreau 1969; Robbins 1978; Oates 1988; Maley 1996).

The Dahomey Gap begins as a narrow strip of open country, first forming the Accra Plains in Ghana then expanding northward and eastward to the border between Nigeria and the Republic of Benin. Although dry forest and savanna species dominate the vegetational composition of the Gap, the region contains an area of moist hill forest along the Ghana-Togo border. Several climatic factors presently maintain the Gap, including low annual rainfall, a bi-annual dry season, and cooling offshore currents combined with the Harmattan Winds that tend to dessicate unprotected areas (Aubréville 1937, 1949; Ern 1988; Oates 1988). The Gap is presently about 400 km wide, but during periods of maximum glaciation it may have been as much as1400 km wide, extending from western Ivory Coast and on to the Cameroon Highlands in the east (Booth 1958a: Booth 1958b; Moreau 1969). On the other hand, Maley (1996) suggests that following the end of the last glacial expansion, forest species may have completely re-colonized the Dahomey Gap.

Several authors have noted the importance of the Dahomey Gap in limiting animal distributions (e.g., Booth 1958a, 1958b; Schiøtz 1967; Moreau 1969), although there have been dissenting views (Robbins 1978). These authors have observed that the areas furthest from the Gap (Liberia and Gabon) contain far more endemic groups than regions closer to the Gap. Moving inward from Liberia and Gabon towards the Gap, habitats become more species-poor. For example, Booth (1958a, 1958b) noted that several species of Primates, Sciuromorpha, Artiodactyla and Hyracoidea occurred in forested areas west and east of the Dahomey Gap but were greatly reduced or absent inside the Gap. In some cases, lineages especially dependent upon lowland rainforest, such as some *Cercopithecus* species, *Procolobus* and *Cercocebus*, are limited in their distribution by the Gap (Booth 1958a, 1958b; Oates 1988). These observations have led these authors to conclude that during periods of maximum glaciation, the then much-larger Dahomey Gap played a more important role as a dispersal barrier than it does at its current width.

Of greater importance, in the opinion of these authors, is the role of rivers in recent western African faunal evolution (Booth 1958a, 1958b; Moreau 1969; Robbins 1978). These authors seem to agree that the Niger River, along with its large delta, and the Volta River are the most important faunal barriers in western Africa, with the Niger being of considerably greater importance. Among primates, the Niger appears to limit *Arctocebus calabarensis*, *Euoticus pallidus*, *Galago alleni*, and has been said to separate *Pan troglodytes verus* from *P. t. troglodytes* (Booth 1958a, 1958b; Grubb 1982; Oates 1988; Grubb 1990; Groves 1993).

However, the Cross and Sanaga Rivers perhaps have had an even greater influence in limiting primate distributions in Nigeria and Cameroon (Oates 1988; Maley 1996). For example, *Cercopithecus sclateri* occurs from the east bank of the Niger River and the eastern portions of the Niger Delta to the west bank of the Cross River in southeastern Nigeria (Oates *et al.* 1992); and *C. erythrotis camerunesis* and *Mandrillus leucophaeus* appear to extend only to the east bank of the Cross. Other primates limited to the east of the Cross River, the Cameroon highlands and the island of Bioko include: *Cercopithecus preussi, C. pogonias pogonias,* two *Procolobus badius* subspecies and *Cercocebus albigena* (Grubb 1982; Oates 1988; Kingdon 1989; Grubb 1990; Groves 1993; Oates 1996). Distributional data for several primates suggest that the Sanaga River also may be an important boundary between *Arctocebus calabarensis/A. aureus, Euoticus pallidus/E. elegantulus Cercopithecus erythrotis/C. cephus, C. nictitans martini/C. n. nictitans. C. pogonias pogonias/C. p. grayi,* and *Mandrillus leucophaeus/M. sphinx*; it also limits the distribution of *Colobus satanas* (Gartlan and Struhsaker 1972; Grubb 1982; Oates 1988; Kingdon 1989; Grubb 1990; Groves 1993; Oates 1996).

The importance of rivers may be over-stated. Oates (1988) suggested that the role of vegetational changes in the vicinity of a river, as well as changes in river course and size over time might reduce or amplify the significance of a given river. For instance, the complex primate distributions near the Cross River may be due to the effects of the Cross River itself, or the result of vegetational changes around the river, including the lack of a large forest block immediately to the west (Oates 1988). Other factors also may influence lineage distribution patterns, such as, inter-specific competition between species with abutting ranges, local differences in forest floristic diversity and relative feeding abundance (Chapman 1983; Oates 1988).

Chimpanzees and forest history

Chimpanzees are behaviorally and morphologically adapted to life in a forested environment (Goodall 1986; Chapman and Wrangham 1993; Groves 1993; Whiten *et al.* 1999); but the extent to which chimpanzees are a strict moist forest taxon is questionable

(Kortlandt 1983; Collins and McGrew 1988; Kortlandt 1991; Moore 1992; Chapman and Wrangham 1993; Fruth and Hohmann 1996; Moore 1996). Although limits to this species' current distribution do not extend far beyond the borders of tropical moist forests (Goodall 1986; Teleki 1989; Oates 1996); they do live in dry forest areas (Kortlandt 1983, 1991; Moore 1992; Moore 1996). However, there is little evidence to suggest that chimpanzees historically have lived outside moist and dry forest zones. Consequently, this species' distribution pattern probably has been influenced by the extent and configuration of tropical forests. Given these observations, it is likely that recent chimpanzee evolution has occurred against a complicated backdrop of multiple refugia, several possible geographic boundaries, shifting vegetation, and the presence of numerous competitive species.

Chimpanzee molecular systematics

It is within this context of multiple Pleistocene refuges and a complex array of geographic barriers across western Africa that the current state of chimpanzee molecular systematics is best understood. Geographically systematic sampling of hairs from chimpanzee sleeping nests has shown that the phylogeography of chimpanzees may be quite different than currently accepted. These studies have featured DNA sequence data from the first hyper-variable region of the chimpanzee mitochondrial genome (HV-I or D-loop). Unlike most primates, male philopatry and female dispersal characterize chimpanzees (Nishida 1979; Pusey 1979; Wrangham 1979; Goodall 1986; Morin 1993). Because mitochondrial DNA is inherited almost exclusively maternally, and is therefore non-recombining, this genome presents an opportunity to track the geographic patterning of chimpanzee genetic diversity (Avise 1994; Hoelzer, Wallman, and Melnick 1998). As such, each new study of different regional chimpanzee populations has used the HV-I locus to build upon previous studies by using the existing sequence data and by adding new data to make increasingly compelling arguments about recent chimpanzee evolution.

The sampling distributions of these previous studies are shown in Figure 1.6. along with geographical terms used frequently throughout this dissertation. Although chimpanzees are typically divided into three lineages that occupy western, central and eastern Africa, these studies have shown that this widely accepted chimpanzee distribution pattern might be more complex. Therefore, in order to clarify the geographical locations of sampling areas and to standardize the geographical terms used throughout this dissertation with previous studies, I will reference samples from these studies using the following geographical divisions:

1) Upper Guinea includes forests west of the Dahomey Gap.
2) Western Nigeria includes forests east of the Dahomey Gap to the Niger River and the western portions of the Niger Delta.
3) Eastern Nigeria and western Cameroon include forests located to the of east of the Niger River and north of the Sanaga River in western Cameroon.
4) Western equatorial Africa includes forests south and east of the Sanaga River in central Cameroon to the Ubangi River and as far south as the Congo River.
5) Eastern Africa includes forests located to the east of the Ubangi River to the western Rift Valley.

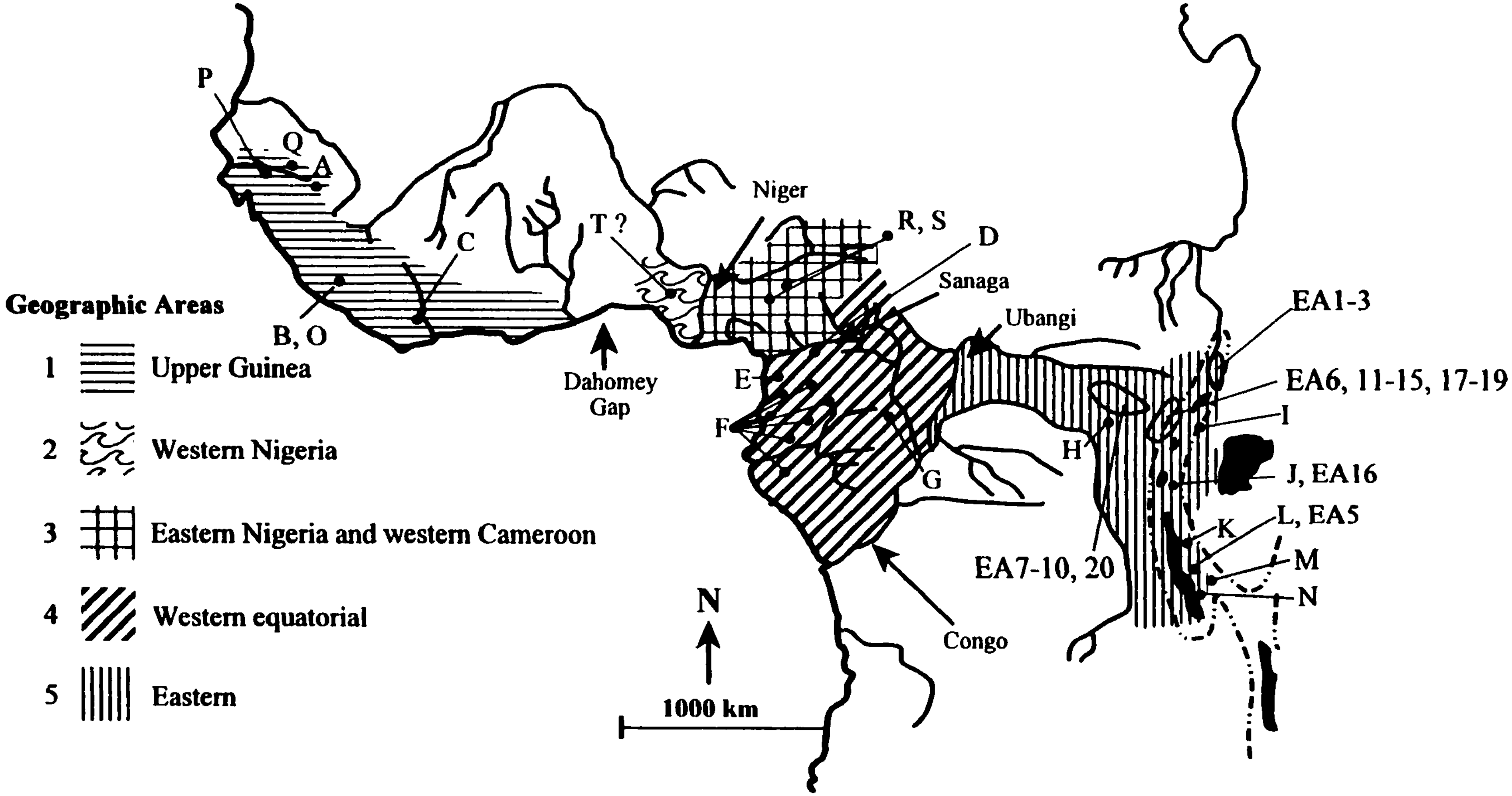


Figure 1.6. Sampling distribution of previous studies and geographical terms used throughout this document. Data from sampling locations A-N taken from Morin (1992). Data from sampling locations EA1-3 and EA5-20 taken from Goldberg (1996). Data from sampling locations R, S and T taken from Gonder *et al.* (1997). Data from sampling locations O, P and Q taken from Gagneux (1998). This study presents new genetic data obtained from chimpanzees in areas 2, 3 and 4 (sampling distribution shown in Figure 2.2). These genetic data are integrated for analysis with samples from other locations listed in this Figure.

The first of these studies (Figure 1.6, sampling locations A-N), published by Morin *et al.* (1994), confirmed the distinct genetic differences between each subspecies found in previous studies of craniometric, pelage and protein variation. The results of this study also revealed that the most genetically diverse of all chimpanzee lineages is *P. t. troglodytes* (Morin 1992; Morin *et al.* 1993; Morin *et al.* 1994). Yet beyond that consensus, the developing scenario became more complicated. Morin *et al.* (1993, 1994) inferred a 1.68 million year isolation of chimpanzees in western Africa from those in western equatorial Africa: an ancient split roughly equivalent to that separating chimpanzees and bonobos. On that basis, they suggested elevating the west African subspecies to full-species status (*Pan verus*).

Morin *et al.*'s results, however, have different interpretations. Jolly *et al.* (1995) argued that these results might be a consequence of recent extinctions or of geographically incomplete sampling. There is a 1,700 km gap in Morin *et al.*'s sampling distribution from western Cameroon to central Ivory Coast (Figure 1.7). Yet, chimpanzees are present in this intermediate area. Based on these observations, Jolly *et al.* (1995) suggested that long term genetic isolation of *P. t. verus* is unlikely. This observation implied that including samples from these intermediate areas might reveal a clinal distribution of genetic variation among chimpanzees in Upper Guinea and those in western equatorial Africa, thus reducing the significant differences between these putative lineages.

Shortly after these papers were published, two investigators reported their findings describing genetic diversity within two chimpanzee subspecies. Goldberg (1996) sampled extensively throughout forests in eastern Africa, including populations

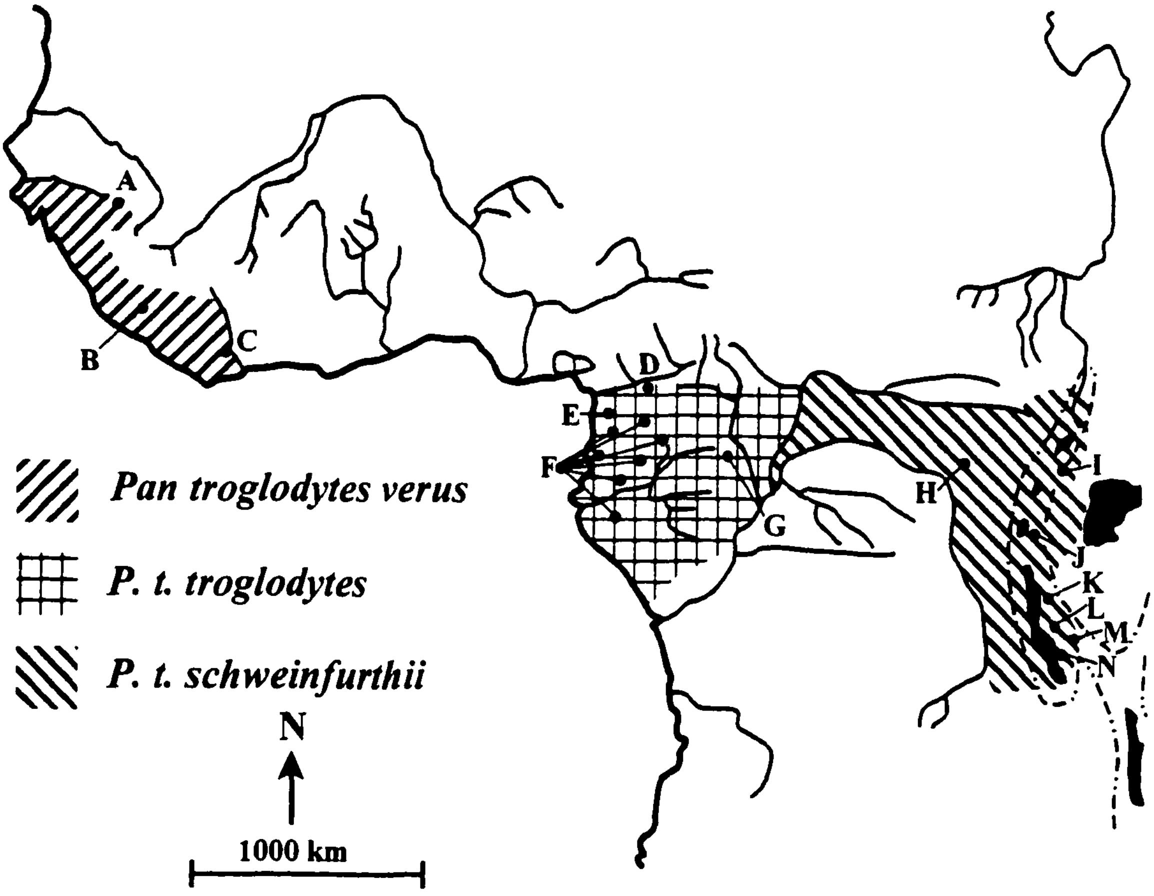


Figure 1.7. Morin *et al.*'s (1994) sampling distribution and proposed geographic distribution of chimpanzees. Note the 1,700 km gap in this sampling distribution from western Cameroon (D) to central Ivory Coast (C).

from the Democratic Republic of Congo (the former Zaire), Rwanda, Tanzania and Uganda (Figure 1.6, sampling locations EA1-EA3 and EA5-EA20). Using 281 HV-I DNA sequences derived from nest hairs, he found evidence of considerable gene flow between populations separated by large distances and by the tributaries of the Congo River. Goldberg (1996, 1998) suggested that genetic diversity within this subspecies was unrelated to distance from refugia, forest area, degree of population isolation or separation by phylogeographic barriers. On that basis, he proposed a recent origin of *P. t. schweinfurthii* and an explosive expansion of these chimpanzees throughout the forests of eastern Africa (Goldberg 1996; Goldberg and Ruvolo 1997; Goldberg 1998).

In a similar genetic survey of chimpanzees in the Upper Guinea forest block, Gagneux (1998) found surprisingly similar results. Samples were collected from chimpanzee communities in Ivory Coast, Mali and Senegal (Figure 1.6, sampling locations O-Q). HV-I sequences, cytochrome-b sequences and genotypes from 11 microsatellite (or STR) loci revealed little genetic division between chimpanzee populations in that region separated by over 1,000 km. These results suggest that this population was panmictic until the very recent past (Gagneux 1998; Gagneux *et al.* 1999). However, deep clades consistently divided HV-I sequences of chimpanzees from the Upper Guinea forest block, but these clades were not concordant with geographic distance or isolation. Further exploration of the characteristics of those deeply branched clades implied that chimpanzee populations in Upper Guinea diversified at an earlier date than those in eastern Africa. In addition, these data suggest that chimpanzee populations in Upper Guinea have experienced a series of migrations of chimpanzees originating further east over a long period of time, unlike the single explosive population expansion that occurred in the forests of eastern Africa (Gagneux, 1998).

Gonder *et al.* (1997) sampled a small number of chimpanzee sleeping nests from eastern Nigeria and skins from western Nigeria (Figure 1.6, sampling locations R, S and T), in an attempt to partially bridge Morin *et al.*'s (1994) sampling gap. We reported a unique HV-I sequence found in the DNA derived from that small sample. We constructed a maximum likelihood tree composed of our sequences combined with Morin *et al.*'s (1994) data set. We identified two distinct lineages of chimpanzee in western Africa that are more different from one another than the central and eastern subspecies are from each other. We concluded that more evidence might suggest a reclassification that would include two forms of chimpanzee in western Africa: one group represented by samples from Upper Guinea, and a group represented by our samples from Nigeria and adjacent parts of Cameroon (Figure 1.8a). We provisionally named this group of chimpanzees from Nigeria and Cameroon, *P. t. vellerosus*, based on the taxon from this area described by Gray in 1862. Additionally, we found evidence of a deep phylogenetic break between chimpanzees in western and western equatorial Africa occurring somewhere in the vicinity of central Cameroon.

Based on these findings, we suggested that the two western groups might be collapsed into a single group, and that western equatorial and eastern chimpanzees might be collapsed into a second group, reflecting a deep west/east division of chimpanzees across Africa (Figure 1.8b). Our findings highlighted themes from earlier work suggesting that high genetic and morphological diversity is found among chimpanzees in

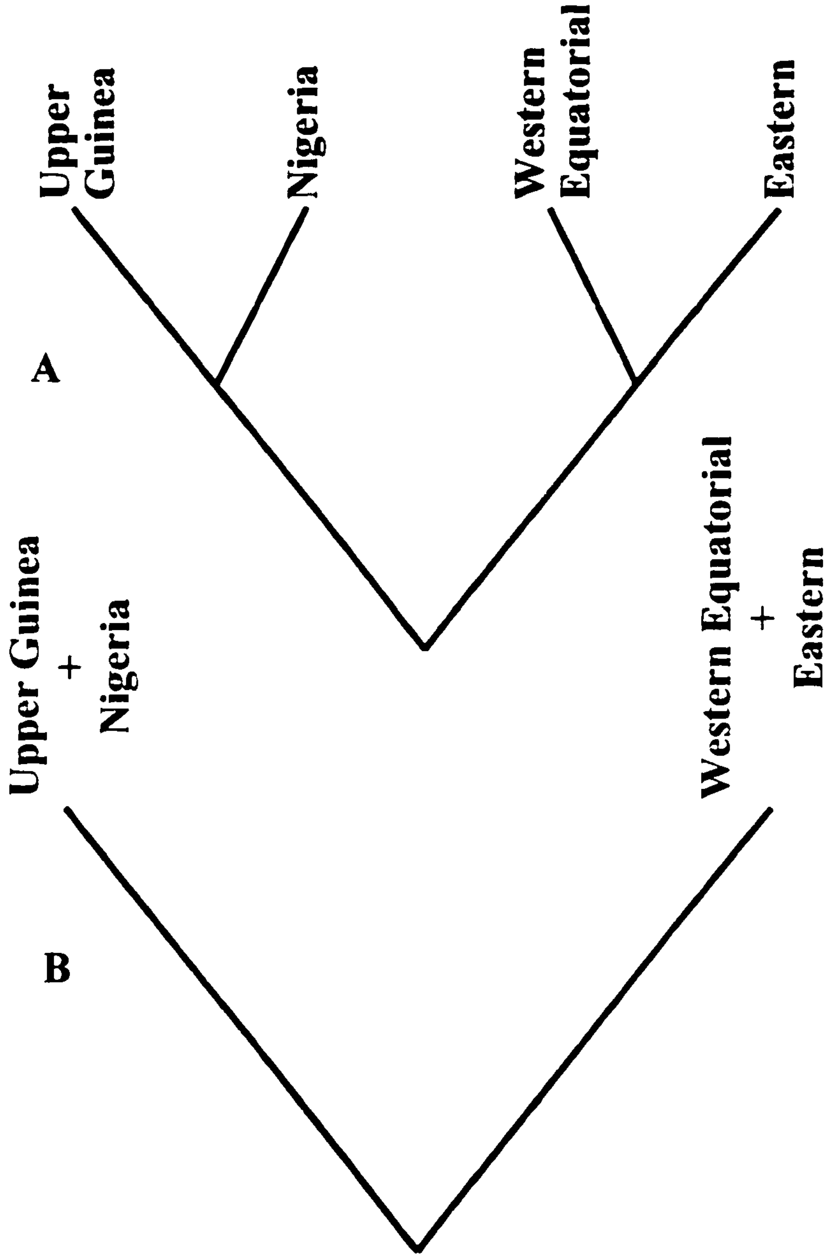


Figure 1.8. Chimpanzee phylogenies proposed by Gonder *et al.* (1997). A) Chimpanzee phylogeny that names chimpanzees with unique HV-I DNA sequence *P. t. vellerosus*. This designation includes chimpanzees from both sides of the Niger River and western Cameroon. B) Chimpanzee phylogeny that collapses chimpanzees in Upper Guinea with those from Nigeria and western Cameroon into one lineage.

western equatorial Africa. The observations imply that one key to understanding recent chimpanzee evolution may lie in this biologically diverse and geographically complex region.

Research questions and structure of the dissertation

Research questions

Previous studies documenting the apportionment of chimpanzee genetic diversity have raised several issues about the phylogeography of this species. These studies suggest that the relationships between chimpanzees probably are more complex than the widely accepted distribution pattern. Foremost among these issues are the phylogenetic relationships between different regional chimpanzee populations. In addition, little is known about the influence of biogeographical barriers on their distribution pattern. This dissertation presents the most geographically comprehensive genetic database available for wild chimpanzees to address these issues. New data obtained from chimpanzees from many localities in Nigeria and Cameroon (Figure 1.6, areas 2, 3 and 4) are integrated with genetic data from previous studies. I use these data to examine the phylogeography of chimpanzees by providing answers to three questions.

1. What is the relationship between chimpanzees in Nigeria and Cameroon and other populations?

Gonder *et al.* (1997) suggested that chimpanzees in Nigeria and western Cameroon might form a separate lineage that is part of a western African group to the exclusion of western equatorial and eastern African chimpanzees. Further evidence of

DNA sequence similarity and monophyly with more westerly populations would provide support for the idea that chimpanzees from Nigeria and western Cameroon are part of a distinct western group. Pronounced DNA sequence dissimilarity with chimpanzees in western equatorial Africa would indicate that the recent evolution of chimpanzees in Nigeria and western Cameroon has occurred without significant contact with chimpanzees in western equatorial African forests.

2. *What geographic barriers, if any, separate chimpanzee populations in Nigeria and Cameroon?*

Gonder *et al.* (1997) suggested two biogeographic hypotheses to explain the relationship between geography and chimpanzee distribution in western and western equatorial Africa. Our evidence did not support the widely accepted separation of western and western equatorial African chimpanzees at the Niger River. Instead, we suggested that the genetic discontinuities lie near the Dahomey Gap to the west and at the Sanaga River to the south.

Evidence of phylogeographic boundaries would include sharp genetic divisions between populations at conspicuous potential dispersal barriers. Genetic data obtained from samples in western Nigeria might provide further evidence of little genetic division between chimpanzees in western and eastern Nigeria, supporting our hypothesis that the Niger River has not played an important role in chimpanzee phylogeography. Instead, more extensive genetic data from chimpanzees in western Nigeria combined with genetic data obtained from chimpanzees in Upper Guinea (Morin 1992; Gagneux 1998) might provide more robust support for our hypothesis that the Dahomey Gap has played an important role in recent chimpanzee evolution.

Our preliminary analyses strongly suggested that a phylogeographic boundary between chimpanzees lay in central Cameroon. This boundary may have had a more pronounced influence on the distribution of this species than either the Dahomey Gap or the Niger River. The Sanaga River limits many other primates in the area. Limits to chimpanzee distribution may lie near the Sanaga River and its tributaries. Evidence that the Sanaga and its tributaries have had a significant influence on chimpanzee distribution patterns would include a sharp division of samples on either side of the Sanaga and its tributaries.

3. How effectively, and for how long, have dispersal barriers limited chimpanzee distribution in Nigeria and Cameroon?

It is difficult to avoid making *post hoc* hypotheses to answer this question because so little is understood about the effects of biogeographic barriers on most primates. Our preliminary data suggested a deep genetic division of western and western equatorial chimpanzees in central Cameroon, and a less pronounced genetic division between chimpanzees separated by the Dahomey Gap. These observations imply that geographic barriers in this region of Africa may have had a substantial and a long term influence on chimpanzee distribution patterns.

Dissertation structure

Chapter 2 details the geographic sampling strategy adopted and the methodology used to collect mtDNA sequence data and microsatellite genotypes. Chapter 3 presents the mtDNA sequences produced during this study, and addresses the first question directly by analyzing several phylogenetic reconstructions of the mtDNA sequence database. The internal nodes of the gene trees produced by the phylogenetic reconstructions are examined more thoroughly in Chapter 4 by various population genetic analyses to explore questions two and three. Chapter 5 presents the microsatellite, or STR, genotype data. Population genetic analyses are presented, and are used to address questions two and three using those data as a nuclear analogue to the mtDNA sequences. Chapter 5 also contrasts the different results found between the two genomes. Chapter 6 synthesizes the conclusions and interpretations presented separately in Chapters 3, 4 and 5. Chapter 6 also presents options for future research on and outlines conservation priorities for chimpanzees in Nigeria and Cameroon.

CHAPTER 2

Materials and Methods

<u>Sampling strategy and tissue collection</u>

Sampling strategy

The range of chimpanzees across southern Nigeria and Cameroon covers a wide array of habitats including montane forest islands, dry forest and savanna mosaics, and large blocks of lowland rainforest. Figure 2.1 illustrates the surface features and habitats found throughout the region. Study locations were chosen to maximize sampling coverage across potential biogeographic barriers including: west and east of the Niger River, north and south of the Cross River, either side of the Cameroon Highlands along a southwest-northeast axis, north and south of the Sanaga River, east of the Mbaum River and into the Congo basin forest expanse near the Dja River (Figure 2.2 and Table 2.1). Samples from captive chimpanzees were collected from the *Pandrillus* Primate Rehabilitation Centers in Calabar, Nigeria and in Limbe, Cameroon. The geographic origin of these captive chimpanzees is uncertain. Consequently, genetic data obtained from those samples were used to optimize laboratory and analytical procedures. Table 2.2 lists samples used for analyses in this study, along with their latitudinal and longitudinal coordinates. Sample name and location abbreviations listed in Tables 2.1 and 2.2 are used throughout this document. Appendix 1 lists all samples gathered for this project.

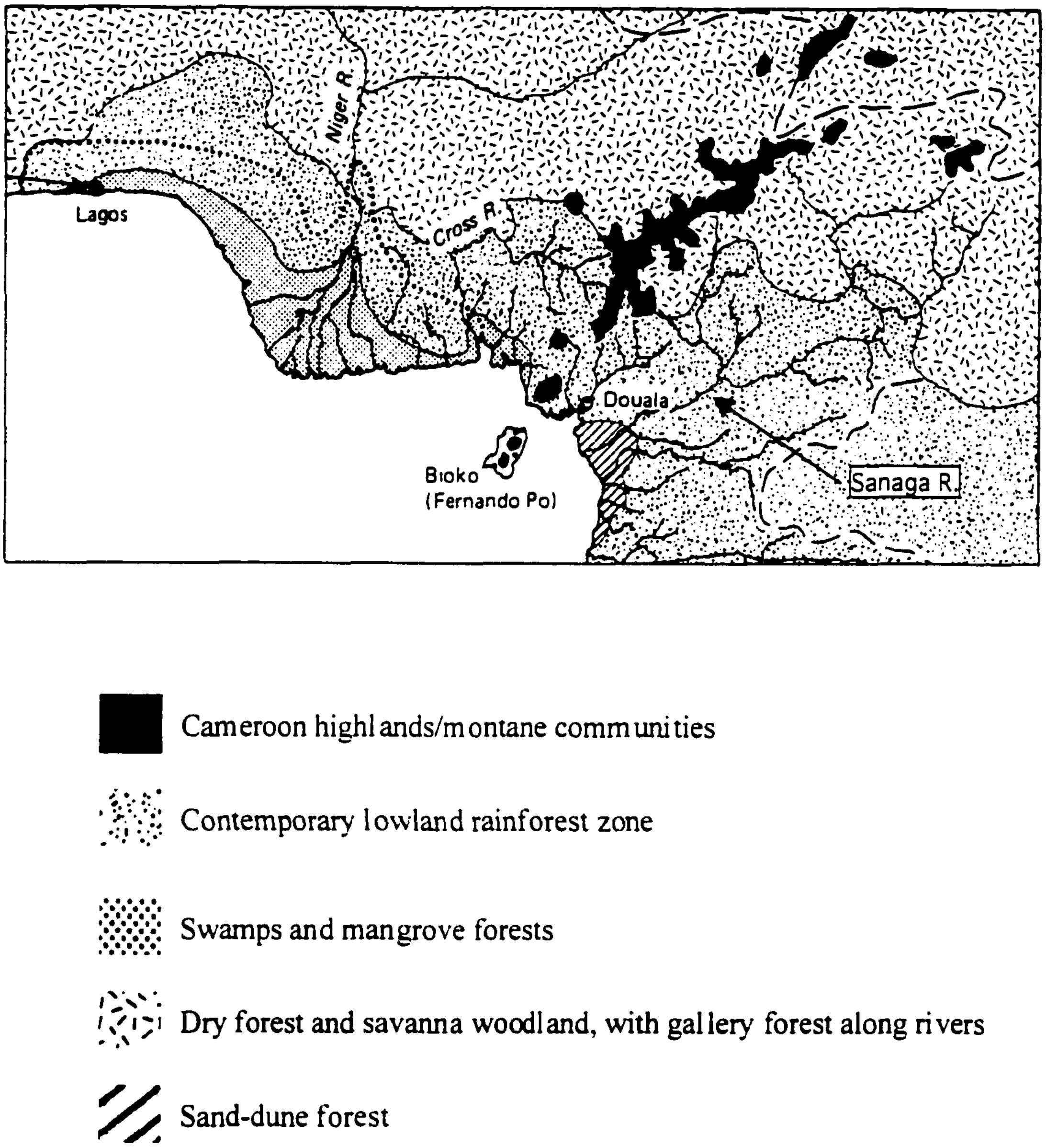


Figure 2.1. Surface features of Nigeria and Cameroon (redrawn from Kingdon, 1989). This map includes a theoretical vegetation cover of the area, but the vegetation cover has been modified by humans.

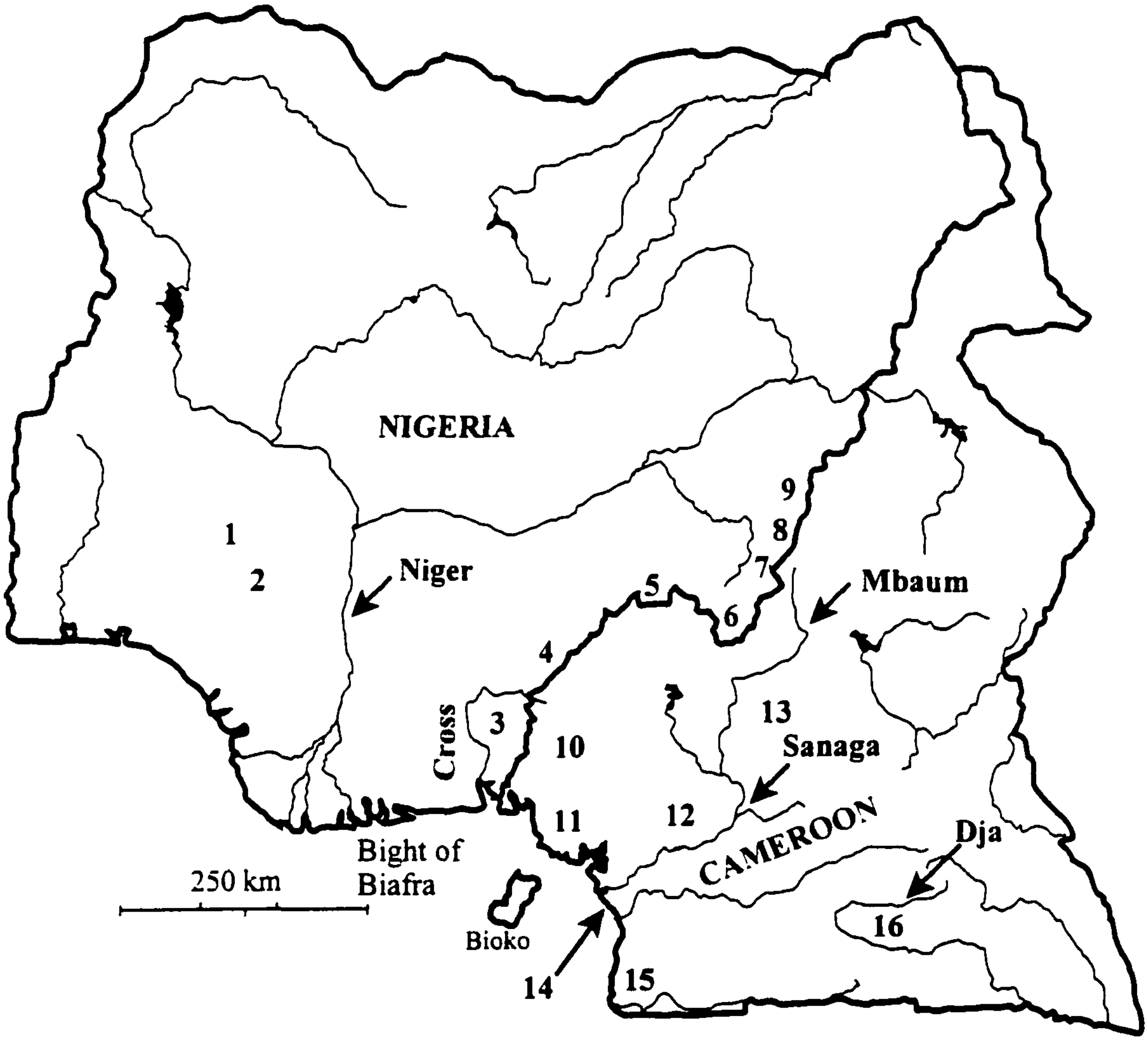


Figure 2.2. Sampling distribution in Nigeria and Cameroon, indicating collecting localities and major rivers. Table 2.1 lists names and descriptions of sampling locations.

Table 2.1 Sampling distribution

Map*	Name	Abbr.	Description
1	Ise Forest Reserve	Ise	lowland moist forest remnant (Agbelusi 1994)
2	Owo Forest Reserve	Owo	lowland moist forest remnant (Agbelusi 1994)
Cross River National Park (3, 4)		CRNP	lowland moist and montane forest (Oates *et al* 1990)
3	-Oban Division		
4	-Okwangwo Division		
5	Akoh Zanto	AKZN	lowland moist forest ascending the slopes of Mambilla Plateau (Chapman, pers. comm.)
6	Ngel Nyaki Forest Reserve	NGNY	montane grassland interspersed with forest along streams and slopes (Dowsett and Dowsett-Lemaire 1989; Barnwell 1991)
Gashaka-Gumti National Park (7, 8, 9)		GGNP	savanna to lowland moist forest mosaic to montane areas (Barnwell 1991; Dunn 1993)
7	-Leinde Fadali		
8	-Kwano		
9	-Gumti		
10	Banyang-Mbo Wildlife Reserve	BYM	lowland moist forest (Faucher, pers. comm.; Prescott *et al.* 1996)
11	Mt. Cameroon	MTCM	lowland moist and montane forests and grassland (Gadsby and Jenkins 1992; Prescott *et al.* 1996)
12	Mosse	Mosse	lowland moist forest (Gadsby and Jenkins, pers. comm.)
13	Manb'ra	Man	savanna-lowland moist forest mosaic (Gadsby and Jenkins, pers. comm.)
14	Douala-Edea Forest Reserve–Tissongo	DEFR	lowland-moist, swamp and sand-dune and mangrove forest (Gartlan and Struhsaker 1972 and Prescott *et al.* 1996)
15	Campo Wildlife Reserve	Campo	heavily-logged lowland moist forest (Mitani 1991; Prescott *et al.* 1996)
16	Dja Biosphere Reserve	DBR	lowland moist forest (Mitani 1990; Prescott *et al.* 1996)

*Map locations shown on Figure 2.2

Table 2.2. List of hair and skin samples, giving sample number, collecting location (with coordinates) and the types DNA analyzed from each sample

*Map locations given in Figure 2.2
†Sample location abbreviations given in Table 2.1
‡ Name of location where the sample was collected
§Coordinates for sample locations were obtained with a Global Positioning System (GPS) unit when possible. Coordinates preceded with the symbol, ≈, were determined by consulting with other researchers or topographical maps.
| |Total number of hairs collected from a sleeping nest or the total number of grams of preserved skin
#Age of sleeping nest determined by the criteria presented in Table 2.3
**An "x" indicates that a complete HV-I sequence was produced for that sample
††An "x" indicates that two or more STR loci were successfully amplified and the alleles scored for that sample

Map*	Sample†	Location‡	Approximate Coordinates§	Hairs||	Nest Age#	HV-I**	STR††
1	ISEFR1	Ise Forest Reserve--Ekiti State	N 07°22.103' E 05°21.869'	9	2	x	x
1	ISEFR3	Ise Forest Reserve --Ekiti State	N 07°22.103' E 05°21.869'	30	3	x	x
1	ISEFR4	Ise Forest Reserve --Ekiti State	N 07°22.103' E 05°21.869'	18	2		x
1	ISEFR5	Ise Forest Reserve --Ekiti State	N 07°22.103' E 05°21.869'	7	1		x
3	CRNP2	Cross River National Park--Oban Hills	≈ N 07°28' E 08°10'	30	2	x	x
3	CRNP3	Cross River National Park --Oban Hills	≈ N 07°28' E 08°10'	14	3	x	x
3	CRNP4	Cross River National Park --Oban Hills	≈ N 07°28' E 08°10'	7	3		x
3	CRNP5	Cross River National Park --Oban Hills	≈ N 07°28' E 08°10'	7	2		x
3	CRNP6	Cross River National Park --Oban Hills	≈ N 07°28' E 08°10'	13	2	x	x
3	CRNP7	Cross River National Park --Oban Hills	≈ N 07°28' E 08°10'	6	2		x
3	CRNP8	Cross River National Park --Oban Hills	≈ N 07°28' E 08°10'	25	2		x
3	CRNP8	Cross River National Park --Oban Hills	≈ N 07°28' E 08°10'	10	2	x	x
3	CRNP10	Cross River National Park --Oban Hills	≈ N 07°28' E 08°10'	12	2	x	
4	CRNP11	Cross River National Park --Okwangwo	N 06°23.8' E 9°19.6'	17	2		x
4	CRNP12	Cross River National Park --Okwangwo	N 06°23.8' E 9°19.6'	15	2	x	x
5	AKZN1	Akoh Zanto	N 06°54.644' E 10°53.993'	3	2		x
5	AKZN2	Akoh Zanto	N 06°54.644' E 10°53.993'	15	2		x

Table 2.2. List of hair and skin samples, giving sample number, collecting location (with coordinates) and the types DNA analyzed from each sample (continued)

Map*	Sample†	Location‡	Approximate Coordinates§	Hairs‖	Nest Age#	HV-I**	STR††
5	AKZN6	Akoh Zanto	N 06°54.644' E 10°53.993'	15	2	x	x
5	AKZN7	Akoh Zanto	N 06°54.644' E 10°53.993'	17	2	x	x
5	AKZN8	Akoh Zanto	N 06°54.644' E 10°53.993'	25	2	x	x
5	AKZN10	Akoh Zanto	N 06°54.644' E 10°53.993'	23	2	x	x
5	AKZN11	Akoh Zanto	N 06°54.644' E 10°53.993'	13	3		x
5	AKZN12	Akoh Zanto	N 06°54.644' E 10°53.993'	32	2	x	x
5	AKZN13	Akoh Zanto	N 06°54.644' E 10°53.993'	3	2	x	x
5	AKZN14	Akoh Zanto	N 06°54.644' E 10°53.993'	48	2	x	x
5	AKZN15	Akoh Zanto	N 06°54.644' E 10°53.993'	14	2	x	
5	AKZN16	Akoh Zanto	N 06°54.644' E 10°53.993'	36	2	x	x
6	NGNY5	Ngel Nyaki Forest Reserve	≈07°04' E 11°04'	4	4	x	
6	NGNY9	Ngel Nyaki Forest Reserve	≈07°04' E 11°04'	18	1		x
6	NGNY10	Ngel Nyaki Forest Reserve	≈07°04' E 11°04'	10	1	x	
6	NGNY14	Ngel Nyaki Forest Reserve	≈07°04' E 11°04'	20	2	x	x
6	NGNY15	Ngel Nyaki Forest Reserve	≈07°04' E 11°04'	400	N/A	x	x
8	GGNP3	Gashaka Gumti National Park--Kwano	N 07°19.733' E 11°35.644'	10	3		x
8	GGNP8	Gashaka Gumti National Park --Kwano	N 07°19.733' E 11°35.644'	20	2	x	x
8	GGNP11	Gashaka Gumti National Park --Kwano	N 07°19.733' E 11°35.644'	23	2	x	x
9	GGNP21	Gashaka Gumti National Park --Gumti	N 07°38.243' E11°44.534'	14	2	x	x
9	GGNP23	Gashaka Gumti National Park --Gumti	N 07°38 243' E11°44.534'	30	2	x	
9	GGNP27	Gashaka Gumti National Park --Gumti	N 07°38.243' E11°44.534'	18	2	x	x
9	GGNP28	Gashaka Gumti National Park --Gumti	N 07°38.243' E11°44.534'	20	2		x
9	GGNP34	Gashaka Gumti National Park --Gumti	N 07°38.243' E11°44.534'	17	2		x
9	GGNP38	Gashaka Gumti National Park --Gumti	N 07°38.243' E11°44.534'	11	2		x
10	BYM1	Banyang-Mbo Reserve	N 05°19.829' E 09°24.904'	16	2	x	x
10	BYM3	Banyang-Mbo Reserve	N 05°19 829' E 09°24 904'	3	3	x	x
10	BYM4	Banyang-Mbo Reserve	N 05°19.829' E 09°24.904'	22	4	x	x

Table 2.2. List of hair and skin samples, giving sample number, collecting location (with coordinates) and the types DNA analyzed from each sample (continued)

Map*	Sample†	Location‡	Approximate Coordinates§	Hairs||	Nest Age#	HV-I**	STR††
10	BYM5	Banyang-Mbo Reserve	N 05°19.829' E 09°24.904'	12	2	x	x
10	BYM6	Banyang-Mbo Reserve	N 05°19.829' E 09°24.904'	15	4		x
11	MTCM2	Mt. Cameroon—Mapanja	N 04°07.362' E 09°09.453'	5	2	x	x
11	MTCM4	Mt. Cameroon—Mapanja	N 04°07.362' E 09°09.453'	40	1	x	x
11	MTCM7	Mt. Cameroon—Mapanja	N 04°07.362' E 09°09.453'	20	1	x	x
11	MTCM8	Mt. Cameroon—Mapanja	N 04°07.362' E 09°09.453'	15	1	x	x
11	MTCM10	Mt. Cameroon—Mapanja	N 04°07.362' E 09°09.453'	73	1		x
11	MTCM11	Mt. Cameroon—Mapanja	N 04°07.362' E 09°09.453'	50	1	x	x
11	MTCM12	Mt. Cameroon—Mapanja	N 04°07.515' E 09°08.626'	52	2	x	x
11	MTCM15	Mt. Cameroon—Mapanja	N 04°07.515' E 09°08.626'	11	1	x	x
11	MTCM16	Mt. Cameroon—Mapanja	N 04°07.515' E 09°08.626'	10	1	x	x
11	MTCM20	Mt. Cameroon—Mapanja	N 04°07.515' E 09°08.626'	30	1		x
11	MTCM21	Mt. Cameroon—Mapanja	N 04°07.515' E 09°08.626'	36	1	x	x
12	MOSSE1	Mosse—North Sanaga	N 04°27.371' E 10°27.130'	3	1	x	x
12	MOSSE2	Mosse—North Sanaga	N 04°27.371' E 10°27.130'	16	1	x	x
12	MOSSE5	Mosse—North Sanaga	N 04°27.371' E 10°27.130'	16	3	x	x
12	MOSSE8	Mosse—North Sanaga	N 04°27.371' E 10°27.130'	22	1		x
12	MOSSE11	Mosse—North Sanaga	N 04°27.371' E 10°27.130'	21	1		x
12	MOSSE12	Mosse—North Sanaga	N 04°27.371' E 10°27.130'	7	2	x	x
12	MOSSE13	Mosse—North Sanaga	N 04°27.371' E 10°27.130'	14	2	x	x
12	MOSSE14	Mosse—North Sanaga	N 04°27.371' E 10°27.130'	30	2	x	x
12	MOSSE15	Mosse—North Sanaga	N 04°27.371' E 10°27.130'	29	2	x	x
12	MOSSE20	Mosse—North Sanaga	N 04°27.371' E 10°27.130'	5	1		x
12	MOSSE21	Mosse—North Sanaga	N 04°27.371' E 10°27.130'	15	1	x	x
13	MAN1	Manb'ra—Ngambe Tikar	N 05°37.281' E 11°40.855'	20	3	x	x
13	MAN3	Manb'ra—Ngambe Tikar	N 05°37.281' E 11°40.855'	32	5	x	x
13	MAN4	Manb'ra—Ngambe Tikar	N 05°37.281' E 11°40.855'	6	5		x

Table 2.2. List of hair and skin samples, giving sample number, collecting location (with coordinates) and the types DNA analyzed from each sample (continued)

Map*	Sample†	Location‡	Approximate Coordinates§	Hairs‖	Nest Age#	HV-I**	STR††
13	MAN6	Manb'ra—Ngambe Tikar	N 05°37.281' E 11°40.855'	23	2	x	x
13	MAN7	Manb'ra—Ngambe Tikar	N 05°37.281' E 11°40.855'	18	5	x	x
13	MAN8	Manb'ra—Ngambe Tikar	N 05°37.281' E 11°40.855'	19	3	x	x
13	MAN10	Manb'ra—Ngambe Tikar	N 05°37.281' E 11°40.855'	3	3	x	x
13	MAN11	Manb'ra—Ngambe Tikar	N 05°37.281' E 11°40.855'	5	3	x	x
13	MAN12	Manb'ra—Ngambe Tikar	N 05°35.936' E11°43.828'	22	2	x	
13	MAN13	Manb'ra—Ngambe Tikar	N 05°35.936' E11°43.828'	62	1	x	x
13	MAN14	Manb'ra—Ngambe Tikar	N 05°35.936' E11°43.828'	10	1	x	x
13	MAN15	Manb'ra—Ngambe Tikar	N 05°35.936' E11°43.828'	10	2	x	
13	MAN16	Manb'ra—Ngambe Tikar	N 05°35.936' E11°43.828'	9	2	x	
14	DEFR1	Douala-Edea Forest Reserve--Tissongo	N 03°35.820' E 09°53.649'	12	2	x	x
14	DEFR2	Douala-Edea Forest Reserve--Tissongo	N 03°35.820' E 09°53.649'	6	3	x	x
14	DEFR3	Douala-Edea Forest Reserve--Tissongo	N 03°35.820' E 09°53.649'	9	2	x	x
14	DEFR4	Douala-Edea Forest Reserve--Tissongo	N 03°35.820' E 09°53.649'	2	1		x
14	DEFR5	Douala-Edea Forest Reserve--Tissongo	N 03°35.820' E 09°53.649'	1	1		x
14	DEFR6	Douala-Edea Forest Reserve--Tissongo	N 03°35.820' E 09°53.649'	16	1	x	x
14	DEFR7	Douala-Edea Forest Reserve--Tissongo	N 03°35.820' E 09°53.649'	7	1	x	x
14	DEFR8	Douala-Edea Forest Reserve--Tissongo	N 03°35.820' E 09°53.649'	26	1	x	x
14	DEFR9	Douala-Edea Forest Reserve--Tissongo	N 03°35.820' E 09°53.649'	3	1	x	x
14	DEFR10	Douala-Edea Forest Reserve--Tissongo	N 03°35.820' E 09°53.649'	6	5		x
15	CAMPO1	Campo Wildlife Reserve	N 02°19.377' E 10°08.401'	11	2	x	x
15	CAMPO2	Campo Wildlife Reserve	N 02°19.377' E 10°08.401'	14	2	x	x
15	CAMPO3	Campo Wildlife Reserve	N 02°19.377' E 10°08.401'	16	4		x
15	CAMPO5	Campo Wildlife Reserve	N 02°19.377' E 10°08.401'	17	4		x
15	CAMPO6	Campo Wildlife Reserve	N 02°19.377' E 10°08.401'	15	1	x	x
15	CAMPO7	Campo Wildlife Reserve	N 02°19.377' E 10°08.401'	11	3		x
15	CAMPO9	Campo Wildlife Reserve	N 02°19.377' E 10°08.401'	45	3	x	x

Table 2.2. List of hair and skin samples, giving sample number, collecting location (with coordinates) and the types DNA analyzed from each sample (continued)

Map*	Sample†	Location‡	Approximate Coordinates§	Hairs| |	Nest Age#	HV-I**	STR††
15	CAMP010	Campo Wildlife Reserve	N 02°19.377' E 10°08.401'	12	3	x	x
15	CAMPO12	Campo Wildlife Reserve	N 02°19.377' E 10°08.401'	40	3	x	x
15	CAMPO13	Campo Wildlife Reserve	N 02°19.377' E 10°08.401'	47	2	x	x
15	CAMPO14	Campo Wildlife Reserve	N 02°19.377' E 10°08.401'	24	2	x	x
15	CAMPO15	Campo Wildlife Reserve	N 02°19.377' E 10°08.401'	27	3	x	x
16	DBR1	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	8	3		x
16	DBR2	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	15	3	x	x
16	DBR3	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	10	3	x	x
16	DBR4	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	2	1		
16	DBR5	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	14	2	x	x
16	DBR6	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	17	1	x	x
16	DBR7	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	15	3		x
16	DBR8	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	25	1	x	x
16	DBR9	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	20	3		
16	DBR10	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	25	1	x	x
16	DBR11	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	6	4		
16	DBR12	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	18	2	x	x
16	DBR13	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	6	2	x	x
16	DBR16	Dja Biosphere Reserve	N 03°11.403' E 12°48.706'	3	1	x	x
	Skin Sample	Suspected Provenance		Quantity| |			
≈1	Ise-sk	ISE, Nigeria		5g		x	x
≈2	Ala-SK	ALA, Nigeria		25g		x	x
≈2	Owo-SK	OWO, Nigeria		15g		x	
Niger Delta	Bomadi-SK	BOMADI—KPAKIEMME, Nigeria		15g			
≈5	Abong-SK	ABONG, Nigeria		20g		x	

All genetic material was obtained non-invasively, from hairs shed in chimpanzee sleeping nests, from hairs plucked from captive individuals and from samples excised from preserved skins. Shed and plucked hairs and preserved skin samples have been proven to be a valuable source of genetic material for studies of wild chimpanzee populations (e.g., Morin 1992; Goldberg 1996; Gonder *et al.* 1997; Gagneux 1998). Although chimpanzees may reuse up to 18% of nests (Plumptre and Reynolds 1997), especially in drier more sparsely vegetated habitats, they typically construct new nests each day at dusk (Fruth and Hohmann 1996). Therefore, hairs collected from a nest are likely to have come from a single individual or a mother-infant pair sharing the nest.

Tissue collection

Samples were collected between April 1997 and March 1998. Sampling locations were targeted either by information from other researchers or by information obtained from local people. Forests were searched at each location for signs of chimpanzees (e.g., vocalizations, sightings) and evidence of recent chimpanzee activity (e.g., feeding sites, new trails, fresh or older nests). If chimpanzees were heard or seen, the animals were tracked until dusk, leaving them in their nests. The following morning, all accessible nests were searched for shed hairs. If evidence of recent chimpanzee activity could not be established in an area but nests were found, all accessible intact nests were searched for shed hair. Between 6 and 22 samples were collected at each location to meet statistical requirements for estimating population isolation and gene flow.

The following information was recorded for each nest in which hair was found: date of collection, height of nest, distance of nest from other nests, relative age of nest, and geographic coordinates. Geographic coordinates were obtained via a Global Positioning System (GPS) unit when possible. If GPS coordinates could not be determined using the GPS unit, the approximate locations of sampling areas were determined by referencing topographical maps or by consulting with other researchers. Nests were assigned relative ages using a 6-point system (Table 2.3).

Table 2.3. Nest aging 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
6 (possibly reused)	Fresher leaves on top of much older leaves

*adapted from Goldberg (1996).

Shed and plucked hairs were collected from nests or from live animals using forceps and sterile latex gloves, taking care to minimize the possibility of contaminating the samples with human tissue. Preserved skin samples were excised using an ethanol- and flame-sterilized razor blade. All samples were placed in glassine envelopes, labeled, and placed in an air- and water tight diver's box with a layer of silica gel desiccant.

Samples were stored at ambient temperatures in the field, and at -20°C after importation to the United States.

Excluding samples obtained from captive animals, the identities of individual chimpanzees are unknown. Consequently, individual chimpanzees may be represented in the sample more than once. Genetic data can be used to unambiguously assign individual identity if unique DNA sequences and genotypes are detected in the sample. All identical DNA sequences from a sampling location were collapsed and treated as a single sample if STR genotypes could not unambiguously define the mtDNA sequences as belonging to separate individuals.

Genetic analysis

The mitochondrial DNA hypervariable region-I (HV-I) and several microsatellite loci were chosen for study. Mitochondrial DNA (mtDNA) has been used extensively to investigate evolutionary relationships among taxa (e.g., Cann, Stoneking, and Wilson 1987; Avise 1989; Vigilant *et al.* 1991; Pesole *et al.* 1992; Avise 1994; Ruvolo 1994; Ruvolo *et al.* 1994; Garner and Ryder 1996; Ruvolo 1996; Pastorini *et al.* 1998). This genome has been used so widely because it is relatively small (≈16 kilobases), generally maternally inherited, haploid, does not undergo recombination and evolves fairly rapidly (Brown 1983; Hoelzer, Wallman, and Melnick 1998; Gray, Burger, and Lang 1999). The D loop, or control region, is a non-coding region that is involved in the replication of the mitochondrial genome (Anderson *et al.* 1981; Pesole *et al.* 1992; Gray, Burger, and Lang 1999; Meyer, Weiss, and von Haesler 1999; Pesole *et al.* 1999). This control region is divided into two hypervariable regions, first identified among human sequences (Kocher and Wilson 1991). The first non-coding hyper-variable region (HV-I) 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; Tamura and Nei 1993), and is therefore an ideal candidate for examining evolutionary relationships between closely related taxa (Simon 1991; Avise 1998). Several studies of genetic diversity in wild chimpanzee populations across Africa have been published, using the HV-I locus (Morin 1992; Goldberg 1996; Gonder *et al.* 1997; Gagneux 1998). My study was designed to allow the direct integration of my results with those from other studies.

However, mtDNA phylogeny may not accurately reflect organismal phylogeny because of differences in male- and female-mediated gene flow, a lower effective number of genes, and random lineage sorting. As such, some authors have argued strongly for nuclear DNA analogues to accompany mtDNA phylogenetic data (Pamilo and Nei 1988; Moritz 1994; Avise and Hamrick 1996; Hoelzer, Wallman, and Melnick 1998). Single tandem repeats (STR's or microsatellites) are rapidly-evolving tandemly-repeated nuclear genome sequences, ranging in size from 150-300bp, whose unit of replication consists of two or more base pair (bp) repeat motifs (Jarne and Lagoda 1996). Like the mtDNA control region, STR's have been used extensively to characterize the genetic structure, demographic history and phylogeography of different populations and lineages (e.g., Bruford and Wayne 1993; Queller, Strassman, and Hughes 1993; Bowcock *et al.* 1994; Jarne and Lagoda 1996; Goldstein and Pollock 1997; Ely *et al.* 1998; Gagneux, Boesch, and Woodruff 1999; Pritchard *et al.* 1999).

STR loci are ideal for these studies because the use of only a few markers has great resolving power for tracking migration routes and examining population geographic isolation (Gagneux, Woodruff, and Boesch 1997; Goldstein *et al.* 1999). If found on different chromosomes, STR's are assumed to be nearly selectively neutral, independent and inherited in a Mendelian fashion. Their short length also makes them amenable to amplifying fragments of degraded DNA, via the Polymerase Chain Reaction (PCR). Processing STR's for analysis is generally far less expensive than generating comparable sequences of nuclear DNA thereby enabling large number of individuals to be screened at a relatively low cost (Bruford and Wayne 1993; Queller, Strassman, and Hughes 1993; Woodruff 1993; Jarne and Lagoda 1996; Bossart and Powell 1998). For these reasons, STR's were chosen for this study to provide a nuclear analogue to the mtDNA sequences. However, Pesole *et al.* (1999) have warned that because the structure, organization and function of the nuclear and mitochondrial genome regions involved in DNA replication are very different, like the control region and some STR's, comparisons between the two genomes may be meaningless.

Laboratory techniques

DNA extraction

DNA was isolated from shed and plucked hair following well-established protocols used in other studies (e.g., Walsh, Metzger, and Higuchi 1991; Morin and Woodruff 1992; Goldberg 1996; Gagneux 1998). DNA from preserved skin samples was obtained using a PUREGENE™ DNA Extraction Kit (Gentra Systems, Inc.).

HV-I amplification and sequencing

HV-I amplification

The polymerase chain reaction (PCR) is a process used to amplify specific DNA sequences many-fold. During the PCR process, a series of thermal cycles each results in the denaturation of a double-stranded target DNA, annealing of oligonucleotide primers to the resulting single strands, and a primer extension period that is catalyzed by a thermostable DNA polymerase. Applications of the PCR technique and typical reaction conditions are well documented (e.g., Innis *et al.* 1990).

Because of the degraded nature of the DNA and low-yield DNA extractions associated with shed and plucked hair, a two stage PCR strategy was used to obtain a template suitable for sequencing. The HV-I locus was amplified in two separate stages: one initial stage in which the entire locus was amplified, and a secondary amplification following post-PCR gel excision of the first PCR product. In the second amplification, either the entire region was re-amplified or the region was re-amplified in two overlapping fragments, dependant upon how degraded the initial PCR template was. Four primers were used for PCR and sequencing (Table 2.4). Figure 2.3 illustrates the amplification and sequencing strategy used for the HV-I locus.

Table 2.4. HV-I PCR and sequencing primers

Primer*	Sequence (5'-3')
L-16041	CTCTGTTCTTTCATGGGGAAGC
L-16111	ATTTCGTACATTACTGCCAGCC
H-16286	GGATGGATTTGACTGTAATGTGC
H-16411	TGTGCGGGATATTGATTTCA

*Primer names indicate the strand to which the primer hybridizes (L = light, H = heavy). Numbers indicate locations of the 3' termini, relative to the human reference sequence (Anderson *et al.* 1981; Goldberg 1996).

Initial PCR reactions were performed in 25μLvolume, using 4μL of genomic DNA extracted from shed and plucked hair or 2μL of genomic DNA from preserved skin. Standard PCR Amplitaq® DNA Polymerase reaction specifications (Perkin-Elmer, Inc.) were used to amplify DNA in a Perkin-Elmer thermocycler set to the following cycling parameters in thin-walled reaction tubes: 94^{O}C for 45 sec, 54^{O}C for 45 sec, 72^{O}C for 1 min, for 35 cycles followed by an indefinite soak at 4^{O}C. The thermocycler was preheated to 94^{O}C prior to loading to reduce non-specific priming and extension during initial ramping.

After initial amplification, samples were combined with 2μL of 6x loading dye, and dried to 10μL. The samples were loaded onto a 1% TAE low-melt agarose gel, and electrophoresed for 45 minutes at 150 volts in TAE buffer. Gels were lightly stained

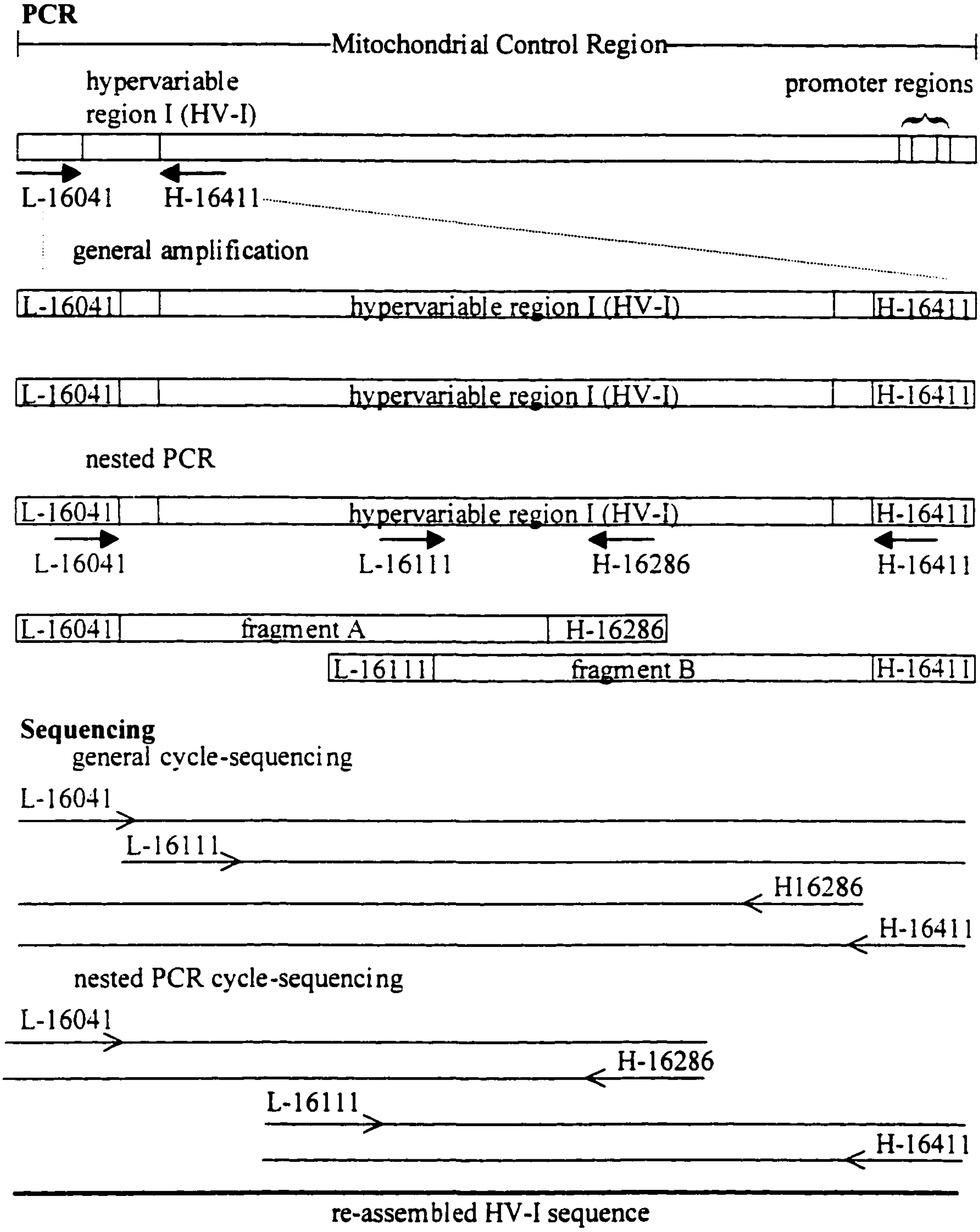


Figure 2.3. Schematic representation of PCR and sequencing strategies. The HV-I locus was amplified from genomic DNA isolated from hair. The resulting product was gel-purified and re-amplified. Depending on DNA quality, the entire region was re-amplified or the second PCR was nested. Each fragment was sequenced twice using each PCR primer on an ABI™ 310 Genetic Analyzer or the 377 Automated Sequencer. The gene sequence was re-assembled by computer from the four resulting electropherograms shown in Figure 2.4.

with ethidium bromide, rinsed in ddH_2O and placed on a UV light table covered with plastic wrap. Using a sterile 1.5mL disposable pipette tips, small "plugs" were removed from the center of the band region in each lane, and were transferred to 1.5mL microcentrifuge tubes containing 500µL ddH_2O. The tubes were heated at $84^{O}C$ for 20 minutes to liberate the gel-bound HV-I fragment prior to re-amplification.

Two 50µL reactions were prepared for each sample using 2µL of the diluted initial product. PCR and thermocycling conditions were identical to the initial amplification, except that 30, instead of 35 cycles were used during the PCR process. Following re-amplification, all samples showing visible single bands on an 8.0% polyacrylimide gel were prepared for DNA sequencing using Qiaquick™ PCR purification kits (Qiagen, Inc.) to remove unincorporated d'NTP's and excess primers. Next, sample quantity and quality were photographed and visually assessed for suitability for DNA sequencing against known quantities of 100bp ladder (Promega Corporation, Inc.).

HV-I sequencing

All sequencing was performed using either the ABI™ 310 Genetic Analyzer or the 377 DNA Sequencer (Applied Biosystems, Inc.). Unlike traditional sequencing methods that rely on radioisotopes and/or toxic chemicals, the ABI™ system relies on laser detection of fluorescently-labeled DNA. In the cycle-sequencing approach to DNA sequencing, a different dye is tagged to each of the four different termination nucleotides. During electrophoresis, a laser scans the product through either a capillary filled with a denaturing gel-like polymer (ABI™ 310 Genetic Analyzer) or a denaturing polyacrylimide gel (ABI Prism® 377 DNA Sequencer). All data were automatically analyzed using ABI™ software and translated into a electropherogram (Figure 2.4a) that displays each base as a different color (Applied Biosystems, Inc., 1993).

Cycle–sequencing reactions were performed with Perkin-Elmer CS™, FS™, d-rhodamine™ or Big-Dye™ terminator cycle-sequencing ready reaction kits (Perkin-Elmer, Inc.). Most samples were processed as quarter reactions (5µL total volume), using the recommended Perkin-Elmer Big-Dye™ cycle-sequencing protocol, adding 2.2µL of purified PCR product. All reactions were processed in a Perkin-Elmer 9600 thermocycler. After cycle-sequencing, reactions were purified using Sephadex Spin Columns (Princeton Separations, Inc.), dried in a DNA Speedy Vac and stored at $-20^{O}C$ to await processing on the ABI™ 310 Genetic Analyzer or ABI Prism® 377.

The fragment was sequenced with each PCR primer, resulting in four reads for each sample (L-16041, L-16111, H16286 and H-1411). For very degraded samples. DNA was amplified using the nested PCR and cycle-sequencing protocols. The fragment amplified with primers L-16041 and H-16286 was sequenced with primers L-16041 and H-16286. The fragment amplified with L-16111 and H-16411 was sequenced with primers L-16111 and H16411. The four sequences produced for each sample were compiled using AutoAssembler® software package provided by ABI™ (Figure 2.4b). For phylogenetic and population genetic analyses, sequences were aligned by visual inspection of the assembled sequences. Previously published sequences obtained from Genbank (e.g., Morin *et al.* 1994; Goldberg 1996; Gonder *et al.* 1997; Gagneux 1998) were aligned with reference to sequences produced for this study. All HV-I locus

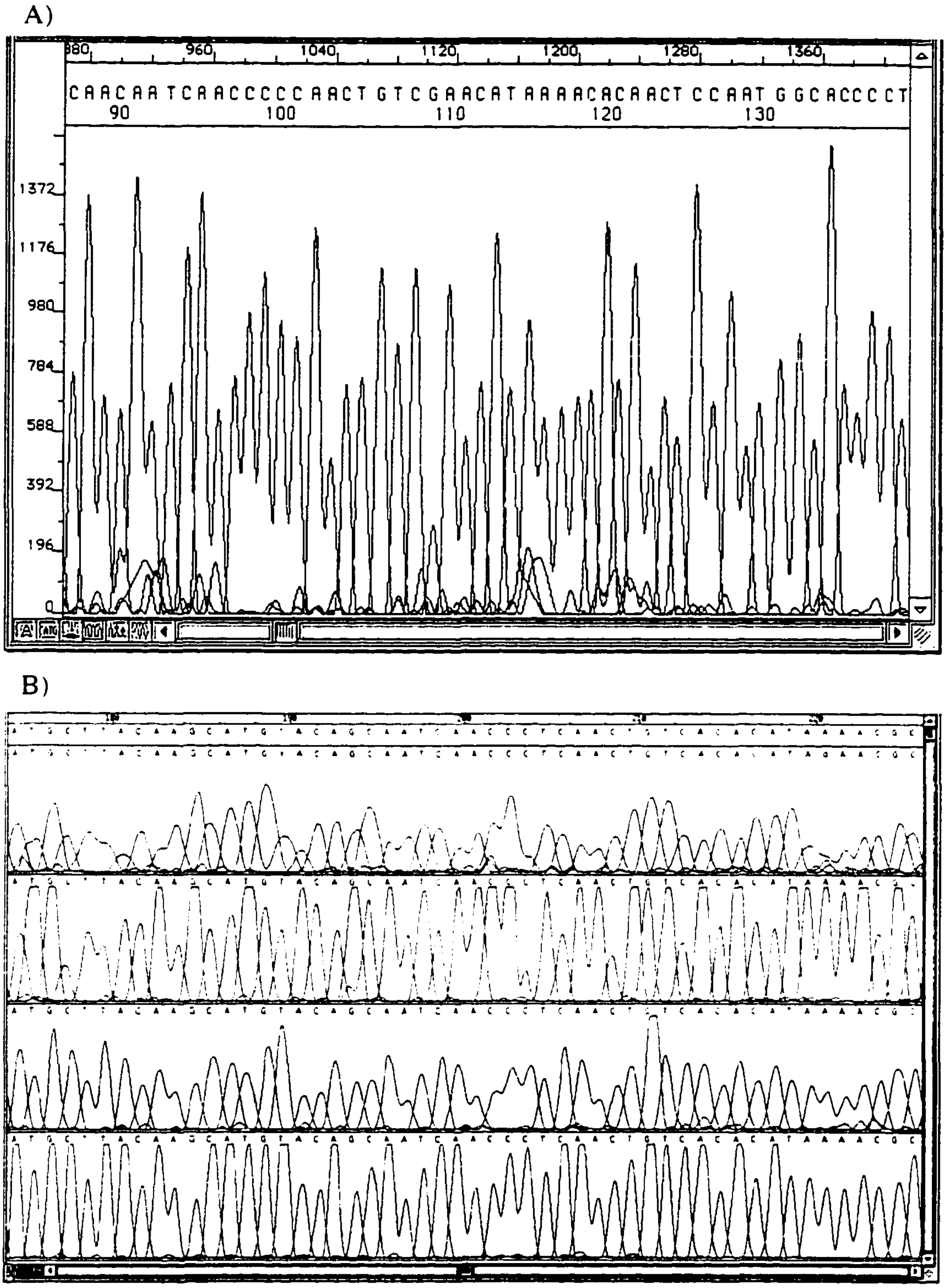


Figure 2.4. Electropherograms produced using the ABI™ sequencing system. The top panel (A) shows an electropherogram produced by the ABI™ 310 Genetic Analyzer. The bottom panel (B) shows four sequences aligned with the Autoassembler® program. Lower case letter denotes a nucleotide ambiguity resolved by visual inspection of the assembled sequence.

phylogenetic and population genetic analytical methods and the results of those analyses are detailed in Chapter 3.

Sequence reliability

Recently, a number of authors have identified mtDNA control region insertions in the nuclear genome of hominoids. Called pseudogenes, these exchanges of genetic information between the mitochondrial and nuclear genomes have occurred consistently and ubiquitously during primate genomic evolution (Fukada, Wakasugi, and Tsuzuki 1985; Nomiyama *et al.* 1985; Collura and Stewart 1995; Zischler, Geisert, and Castersana 1998; B. Haahn unpublished data). Non-specific PCR primers may inadvertently anneal to these sites during the PCR process, amplifying the region of interest and/or the pseudogene contaminant. If a pseudogene was incorporated into a phylogenetic analysis, the separately evolving nuclear sequence would not be homologous to the mitochondrial sequence, thereby confounding all results and interpretations of the analysis.

It is unlikely that pseudogene HV-I sequences were produced during this study. Nucleotide positions that are ambiguous and unresolvable in multiple reads from one sample indicate the presence of pseudogenes due to the presence of more than one sequence in a given reaction. However, my quadruplicate sequencing reads suggested that only a single sequence was present in the assembled reads. Additionally, sequences generated during this study were aligned for analysis with sequences generated by four independent laboratories across the world. The sequence alignments suggest that either each laboratory has sequenced the same or similar HV-I pseudogene or that each laboratory has published actual HV-I sequence. Finally, in any case where sequence ambiguity was detected, fresh extracts, PCR product and cycle-sequencing reactions were prepared. In each case, the second round of sequencing produced more reliable and consistent results, suggesting either degraded DNA or technical error as the source of the ambiguity.

The HV-I sequences from some samples are very different from the HV-I sequences obtained from chimpanzees in the same location (i.e., CRNP6, CRNP12, BYM3, Campo9). For each sample, I completed a second round of DNA isolation, DNA amplification via PCR and cycle-sequencing. In all cases, the first and second rounds of sample processing produced identical HV-I sequences.

STR genotyping

STR fragment amplification and detection

Due to the sensitivity of STR fragment amplification, all DNA extracts were purified using a GENECLEAN® II kit (BIO 101®, Inc.) to remove protein contaminants and small DNA fragments. Protein contaminants interfere with product amplification, and small DNA fragments can cause mis-priming during PCR, leading to inaccurate allele scoring. Samples were successfully amplified in a 10μL volume under the following conditions: 10mM Tris-HCL pH 8.3, 50mM KCL, 0.6 mM $MgCl_2$, 1.5 pmol of each primer, and 0.4 units of Perkin-Elmer AmpliTaq™ Gold DNA Polymerase. In order to further reduce mis-priming problems, reactions were processed in a either a Perkin-Elmer 2400 or 9600 thermocycler programmed for a touchdown PCR protocol for 20 cycles ending at $48^{O}C$, with a final 25 cycles at $48^{O}C$ (Don *et al.* 1991; Rithidech, Dunn, and Gordon 1997; Morin *et al.* 1998).

Several human STR primers were examined for variation among the chimpanzee samples, including di-, tri- and tetra-nucleotide repeats. Loci with more complex motif patterns were examined as well. Table 2.5 summarizes information for all loci screened for this study. Alleles for each sample were assessed using either the ABI™ 310 Genetic Analyzer or the ABI Prism® 377 DNA Sequencer. In this system, fluorescently-labeled STR primers are incorporated along the flanking regions of the amplified fragment. Similar to automated DNA sequencing using this system, amplified product is scanned by laser during electrophoresis using either a capillary filled with a denaturing gel-like polymer (ABI™ 310) or a denaturing polyacrylimide gel (ABI™ Prism® 377; Long Ranger Singel™ Packs, FMC Bioproducts Corp.). Three different dye colors are available, facilitating detection of at least three different loci at a time. A fourth dye is reserved for the DNA size standard that enables the accompanying ABI™ software to accurately determine the size of each fragment.

Gagneux, Boesch and Woodruff (1997) demonstrated that allelic drop out (amplification of only one allele in a true heterozygote) occurred in 31% of single shed chimpanzee hair STR amplifications. This problem is important because it can lead to assigning inaccurate genotypes and an under-estimate of locus heterozygosity. Multiple amplifications from multiple extractions for hairs obtained from each sleeping nest were suggested to minimize this problem (Gagneux, Boesch, and Woodruff 1997; Goosens, Waits, and Taberlet 1998). In order to minimize the likelihood of allelic drop out, two to ten extractions were performed on each nest hair collection. Following these multiple extractions, PCR reactions were performed separately using DNA from each extract; or extracts were combined prior to GENECLEAN® II purification. Whenever possible, all

Table 2.5. STR loci

Marker*	Locus	Size (bp)	Repeat Unit	Primer Sequence (5')	Primer Sequence (3')
Mfd 3	APOA2 (D1)	120-162	AC	GGTCTGGAAGTACTGAGAAA	GATTCACTGCTGTGGACCCA
Mfd23	D16S265	78-104	AC	AGTCCTCTGTGCCACTTTGT	CCAGACATGGCAGTCTCTA
HumPla2a	HumPla2a	65-98	AAT	GGTTGTAAGCTCCATGAGGTTTAGA	GTCCTAGGAGCTAGAGATACAGC
D4S1652	G00-228-893	114-162	ATCT	AATCCCTGGGTACATTATATTTG	CAGACATTCTTTATTCTTTACCTCC
D7S1809	G00-364-834	202-242	AGGA	AGGCAAGAGCAGTAGCAAGA	TCCACTTTAAATCAGCAGCC
D9S303	G00-218-317	142-194	GATAxCAGA	CAACAAAGCAAGATCCCTTC	TAGGTACTTGGAAACTCTTGGC
D11S1984	G00-364-803	104-196	CAAA	GGGTGACAGAGCAAAATTCT	ACACCTGGATCTTGGACTCA
D13S317	CHLC.GAT A7G10.415	155-255	TATCxATCT	ACAGAAGTCTGGGATGTGGA	GCCCAAAAAGACAGACAGAA
D16S539	G00-228-700	152-168	ACAGxGATA	GATCCCAAGCTCTTCCTCTT	ACGTTTGTGTGTGCATCTGT
D20S470	G00-364-824	167-303	TTCCxCCTTx CCTTxTC	CCTTGGGGGATATAGCCTAA	TGAGTGACAGAGTGATACCATG

*Marker names reflect chromosome number and location.

amplified samples for a given locus were confirmed by two independent amplifications. Homozygous samples were confirmed by a third independent amplification. Conflicting allele sizes for all samples were either re-processed until the ambiguities were resolved or were removed from the study. Errors in allele size estimation do occur when switching between the ABI™ 310 and ABI Prism® 377 systems. However, allele sizes between machines do not vary more than +/- 2 bp (ABI™ GeneScan® Reference Manual). Positive controls of known size were run alongside each sample set to diminish variable results between machines.

Allele scoring

Allele sizes were determined by combining ABI™ software analysis (GeneScan® and Genotyper®) and discrimination by visual inspection (Figure 2.5). Allele sizes were considered genuine when they were within a specified range of values closely correlated with published allele sizes for each locus. For instance, alleles with tetra-repeat motifs were scored within a range of +/- 2bp of published allele sizes. Newly identified allele sizes for each locus were confirmed by multiple runs of each unique allele. Allele sizes for each sample were entered into a spreadsheet and prepared for population genetic analyses. All population genetic analytical methods used and the results of those analyses for all STR loci are detailed in Chapter 4.

STR reliability

In addition to the problems with allelic dropout (Gagneux, Boesch, and Woodruff 1997; Goosens, Waits, and Taberlet 1998), a growing body of evidence suggests that

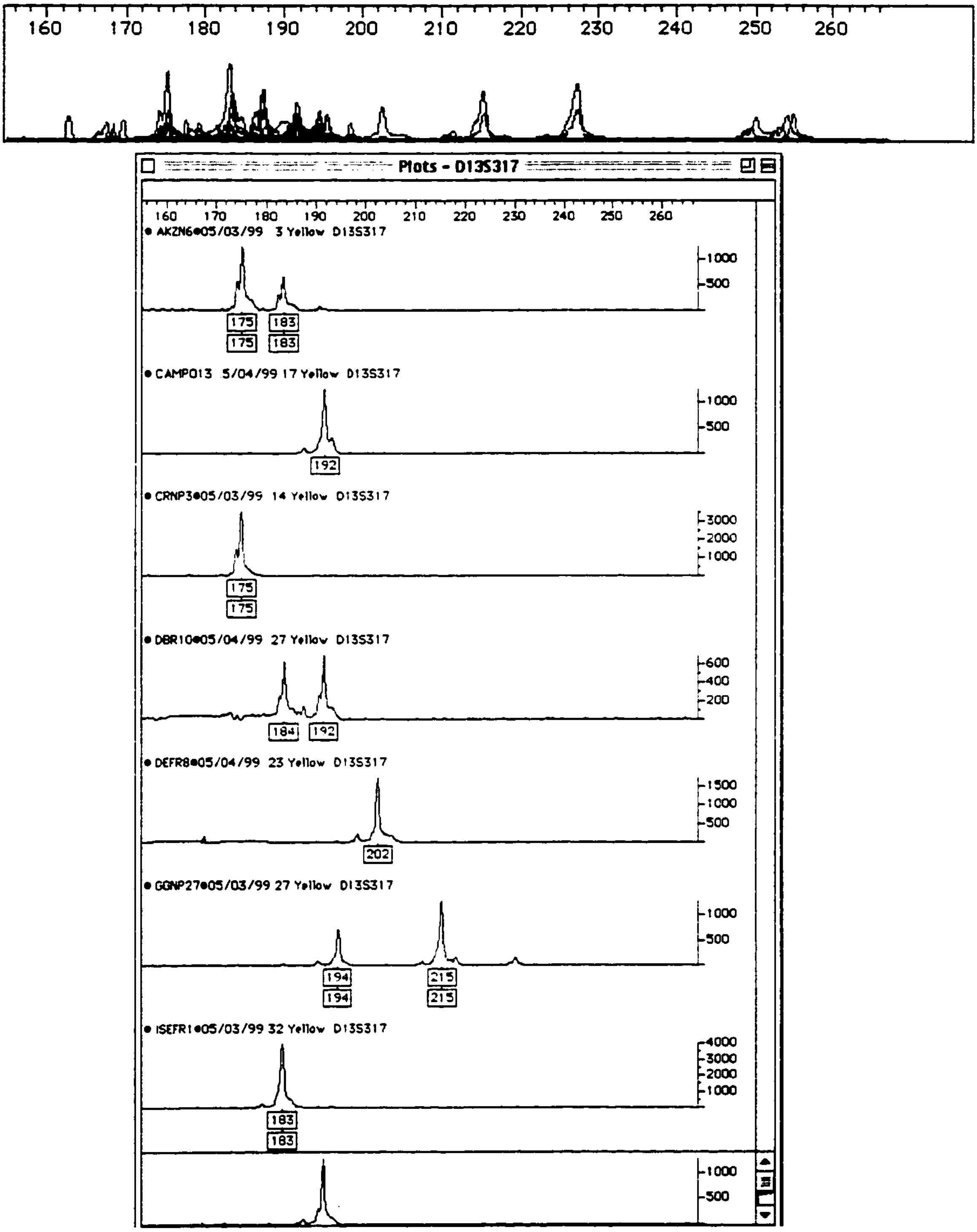


Figure 2.5. Allele sizes analyzed using ABI™ Genotyper® software. The top panel shows alleles for all samples at the D13S317 locus. The bottom panel shows alleles for selected samples.

STR data should be interpreted carefully for other reasons. STR loci mutate rapidly with an estimated rate of 10^{-2} to 10^{-5} per generation (Chakraborty *et al.* 1997; Schug *et al.* 1998). STR allele size variation is not well understood, but is believed to follow a step-wise mutation model in which alleles differ by a certain number of repeat units (Bowcock *et al.* 1994; Goldstein *et al.* 1995). Moreover, there appears to be a positive asymmetry in the mutation process, such that one or two repeat units may be added to an allele each generation, leading to an increase in allele size over time (Goldstein *et al.* 1995; Rubinsztein *et al.* 1995; Slatkin 1995).

However, several authors have shown that differences in allele length are not always attributable to simple changes in repeat number (Amos and Rubinstzein 1996; Amos *et al.* 1996; Jin *et al.* 1996; Goldstein and Pollock 1997; Brinkmann *et al.* 1998). There seem to be constraints on the maximum number of repeats a given allele can possess (Bowcock *et al.* 1994; Garza, Slatkin, and Freimer 1995; Goldstein and Pollock 1997). In comparisons between some lineages, constraints on maximum allele size are confounded further by degradations in the step-wise mutation process over time that can be coupled with a disruption in the purity in the series of repeat units (Crouau-Roy *et al.* 1996; Garza and Freimer 1996; Goldstein and Pollock 1997). Moreover, the series of repeat units makes up only small portion of the amplified fragment. Insertions and deletions outside the repeat unit array but contained within the flanking regions, can further complicate accurate allele identification and scoring (Grimaldi and Crouau-Roy 1997; McAllister and Werren 1999). Selective forces may also operate on certain STR loci. A number of heritable diseases have been associated with some tri-nucleotide STR loci, such as, Fragile-X, Mytonic Dystrophy and Huntington's Disease (Jarne and Lagoda 1996; Chakraborty *et al.* 1997; Goldstein and Pollock 1997).

These problems are evident in comparisons of the same STR loci between lineages across broad time scales, and are primarily attributed to homoplasy in STR allele sizes and a limited understanding of how the mutation process varies between STR loci and between lineages (Schug *et al.* 1998; Taylor, Sanny, and Breden 1999). Most STR loci used to examine genetic diversity are not species-specific. Heterologous markers have been used extensively to characterize genetic diversity among a wide array of taxa under the assumption that the flanking regions and repeat arrays of these loci must be highly conserved because the primers of these markers amplify DNA in different lineages, (e.g., Moore *et al.* 1991). For example, all STR primers used to amplify chimpanzee DNA for this study are human-derived primers. However, non-specificity makes it difficult to determine not only if the correct region has been amplified, but also whether the locus is under comparable constraints in the target taxa. Without directly sequencing the PCR product there is no way to be certain the differences in the number of repeat units are the source of fragment size variation (Garza and Freimer 1996).

In addition, there appears to be an ascertainment bias in heterologous markers (Blanquer-Maumont and Crouau-Roy 1995; Garza, Slatkin, and Freimer 1995; Crouau-Roy *et al.* 1996; Garza and Freimer 1996). That is, the alleles seem to be much longer and more polymorphic in the taxon for which the marker was identified than in other taxa (Goldstein and Pollock 1997). The ascertainment bias hypothesis has been used to explain interspecific allele size differences between humans, chimpanzees, baboons and macaques, making these markers suspect in their resolving power for reconstructing historical relationships between primate lineages across different time scales (Blanquer-Maumont and Crouau-Roy 1995; Garza, Slatkin, and Freimer 1995; Rubinsztein, Leggo, and Amos 1995; Garza and Freimer 1996). More recent evidence suggests, however, that ascertainment bias is not the sole explanation for interspecific allele size variation; and that the position and size of STR alleles is generally conserved between these lineages (Domingo-Roura *et al.* 1997; Morin, Kanthaswamy, and Smith 1997; Watanabe *et al.* 1997; Cooper, Rubinsztein, and Amos 1998; Crawford *et al.* 1998; Morin *et al.* 1998).

However, even among very closely related lineages, homoplasy and differences in STR mutation processes sometimes can confound accurate reconstructions of the relationships between populations and lineages. Estoup *et al.* (1995), for example, found high levels of allelic homoplasy that resulted in an underestimate of the inferred population and lineage divergence between honeybee subspecies. Similar results have been reported across a broad range of taxa including, plants (Doyle *et al.* 1998), invertebrates (Jin *et al.* 1996; Metzgar *et al.* 1998; Viard *et al.* 1998), fish (Zardoya *et al.* 1996; Jones *et al.* 1999; Taylor, Sanny, and Breden 1999), birds (Primmer and Ellegren 1998) and mammals (Blanquer-Maumont and Crouau-Roy 1995; Rubinsztein, Leggo, and Amos 1995; Garza and Freimer 1996).

Allelic homoplasy, varying mutational processes and violations of selective neutrality assumptions at STR loci may, therefore, lead to inaccurate estimates of within-population genetic diversity, among-population genetic divergence and reconstructions of phylogenetic relationships (Taylor, Sanny, and Breden 1999). Consequently, STR's may be of little value in reconstructing population histories even if demographic events of interest occurred fairly recently. Many of the problems with STR's have only been detected over the past five to six years. There may be many more undiscovered problems. For applications other than pedigree analysis and paternity assessment, information obtained from using STR loci should be regarded cautiously.

Despite the problems with STR loci, they may be useful for assessing the apportionment of genetic diversity among chimpanzees. STR's provide a fairly rapid and relatively comprehensive means of assessing nuclear DNA variability (Jarne and Lagoda 1996). In addition, using these loci also may indicate other types of genetic screening that might be more appropriate for testing very specific hypotheses about lineage relationships and population histories (Goldstein and Pollock 1997; Bossart and Powell 1998). Given that little is known about nuclear DNA variation among chimpanzees, the relative ease of amplifying STR loci from degraded DNA template and the different time scales represented by different types of STR loci, STR's may be valuable for an initial evaluation of the apportionment of genetic diversity among chimpanzees near putative biogeographic boundaries.

CHAPTER 3

Phylogenetics of the mtDNA Hyper-Variable Region-I of Chimpanzees in Western Africa

Introduction

A maximum of 254 unique and complete chimpanzee HV-I mtDNA sequences were used in the analyses presented in this chapter. Although over 150 HV-I sequences were produced during this study, only 79 sequences obtained from chimpanzees in Nigeria and Cameroon were used for analysis. These HV-I sequences represent chimpanzees across the study area. Most of the sequences were derived from the shed hairs collected at the sampling locations depicted in Figure 2.2, and a few sequences were obtained from preserved skin samples (Table 2.2). These sequences were combined for analysis with HV-I sequences from previous studies. From Morin's (1992) pan-African sample, 25 unique sequences were used. From Goldberg's (1996) sample of chimpanzees in eastern Africa, 119 unique sequences were used. From Gagneux's (1998) sample of chimpanzees from Upper Guinea, 31 unique sequences were used. Seventy-one HV-I sequences produced during this study and several sequences from these previous studies were excluded from analyses due to uncertain geographic origin, missing data (more than 10 consecutive missing base pairs [bp]) or were grouped together because their nucleotide composition is identical (i.e., composite haplotypes).

Some of the unique sequences are composite haplotypes. All composite haplotypes were constructed using the computer program MacClade (Maddison and Maddison 1996). Sequences were collapsed into a composite haplotype if they were identical or if the resolution of missing or ambiguous data could make them identical. Ten sequences from chimpanzees in Nigeria and Cameron were shared either between or within sampling locations, and were collapsed into three unique composite haplotypes. The first three letters of the author's last name, the abbreviation COMP, and a numerical designation name each composite haplotype (e.g., GONCOMP01).

The 254 sequences used in final analyses were 346bp in length and yielded 201 polymorphic sites. Appendix 2 lists the 254 aligned sequences used for analysis during this study along with human and bonobo sequences that were used to root gene trees. The human and bonobo HV-I sequences were chosen at random from the Genbank database. The end of Appendix 2 lists the sequences of the composite haplotypes from Nigeria and Cameroon used for this study.

In this chapter, I present several phylogenetic reconstructions based on the entire HV-I sequence database. These reconstructions are used to infer overall aspects of recent chimpanzee evolution throughout Africa, and specifically are used to examine the relationships between chimpanzees in Nigeria and Cameroon and chimpanzee populations elsewhere. This chapter provides an initial assessment of the distribution of genetic diversity at the HV-I locus, and attempts to assess the most likely phylogeographic barriers influencing chimpanzee distribution. The first section of this chapter addresses the validity of selective neutrality assumptions at the HV-I locus. The second section of this chapter presents phylogenetic reconstructions of the HV-I database.

The third section of this chapter compares molecular evolution at the HV-I locus in chimpanzees and humans and addresses how different types substitutions may influence phylogenetic reconstruction. Finally, relative divergence dates are presented for the chimpanzees lineages inferred from the phylogenetic reconstructions. Chapter 3 concludes with a brief summary of these analyses.

Selective neutrality

Theory and methods

The mitochondrial DNA control region is a mainstay of studies concerned with microevolutionary processes and phylogeography (Avise 1998). These studies rely on the notion that neutral processes, like genetic drift, determine the degree of genetic differentiation between populations and lineages. The absence of selective forces influencing the mtDNA control region is a key assumption of analyses used to detect and to diagnose microevolutionary processes (Avise 1989, 1994; Hoelzer, Wallman, and Melnick 1998). However, the assumption that the control region is not subject to selection is controversial. The control region contains the main regulatory elements for mtDNA transcription and replication. It is the binding site for numerous molecules such as DNA and RNA polymerases and other transcription and regulatory factors (Saccone, Pesole, and Sbisa 1991; Gray, Burger, and Lang 1999). Consequently, the control region may be subject to functional and evolutionary constraints (Meyer, Weiss, and von Haesler 1999). In addition, some phylogenetic analyses (Hey 1997; Harris and Hey 1999) and statistical tests of neutrality (Merriwether *et al.* 1991; Aris-Brosou and Excoffier 1996) suggest that selective forces have influenced the human mtDNA control region. The interpretations of the phylogenetic reconstructions presented in this chapter and the population genetic analyses presented in Chapter 4 rely on the assumption that the HV-I locus is not subject to selective forces (Avise 1994, 1998; Hoelzer, Wallman, and Melnick 1998). Consequently, it is important at the outset to test the assumption that data from this genetic region indeed conform to a model of neutral mutation.

Tajima (1989) described a statistical test of neutral mutation, derived from Kimura's (1969) infinite sites model. Tajima's test compares two estimators of the population parameter, θ. The first estimator is based on the number of polymorphic sites in the sample ($\hat{\theta}\pi$). The second estimator is based on the mean number of pairwise differences between mtDNA haplotypes ($\hat{\theta}s$). Both estimators should be roughly equal under the infinite-sites model. Tajima's test statistic, *D*, is defined as:

$$3.1) \qquad D = \frac{\hat{\theta}\pi - \hat{\theta}s}{\sqrt[var]{(\hat{\theta}\pi - \hat{\theta}s)}}$$

where $\hat{\theta} = \pi$ and $\hat{\theta}s = S/\sum_{i=0}^{n-1}(1/i)$ and s represents the number of segregating sites in the sample. The significance of the test statistic, *D*, is determined by a beta-distribution limited by a distribution of possible *D* values determined from a simulated population at equilibrium (Tajima 1989, p. 589). Significant departures from the expected distribution of *D* values indicates a rejection of the null model of selective neutrality, and might point to selection, recent demographic change or mutation rate heterogeneity between sites influencing genetic divergence between individuals and populations (Tajima 1989).

Results and interpretation

Results from Tajima's test were calculated using the Arlequin software package (Schneider *et al.* 1997), and are presented in Table 3.1. The results include *D* statistics calculated for each sampling location in Nigeria and Cameroon, for possible regional divisions within that data set and for chimpanzee sequences from other studies. Significant departures from neutral expectations were not detected at any sampling location or within any region. These results are surprising given the significant departures from null expectations detected in other analyses presented in this chapter and in Chapter 4. Nevertheless, the results from Tajima's test validate the use of models of interpopulational differentiation, which assume the predominance of selectively neutral evolutionary forces.

Phylogenetics

Character- and algorithm-based methods are both powerful tools for phylogenetic reconstruction. The cladistic method is a character-based approach that generally employs the principle of maximum parsimony (Hennig 1966). Parsimony methods operate by selecting trees that minimize the total tree length required to explain a given set of data, making it possible to compare optimal and sub-optimal gene trees (Swofford *et al.* 1996). Simulated and empirical analyses have demonstrated that the cladistic approach reliably approximates phylogenetic relationships among taxa (e.g., Nei 1987; Saitou and Nei 1987; Miyamoto and Cracraft 1991; Hillis *et al.* 1992).

The maximum likelihood approach is also a character-based method. Initially developed by Cavalli-Sforza and Edwards (1967), the maximum likelihood approach

Table 3.1. Tajima's neutrality test applied to sampling locations and regional populations

Location*	n	S†	Pi‡	D§	P[a]
Ise	2	n/a	n/a	n/a	n/a
CRNP	6	51	26.67	1.24	ns
AKZN	8	32	12.18	-0.07	ns
NGNY	3	n/a	n/a	n/a	n/a
GGNP	8	33	13.11	0.16	ns
BYM	4	n/a	n/a	n/a	n/a
MTCM	9	34	13.89	0.56	ns
Mosse	9	33	13.36	0.51	ns
Man	10	61	23.98	0.55	ns
DEFR	7	42	15.29	-0.62	ns
Campo	10	37	11.24	-0.68	ns
DBR	10	64	21.82	-0.17	ns
Nigeria and western Cameroon[1]	49	120	18.07	-1.18	ns
Western Equatorial Africa[1]	37	107	25.86	0.03	ns
Total[2]	86	151	26.75	-0.37	ns
Upper Guinea[3]	78	55	12.10	0.27	ns
Eastern Africa[4]	123	94	8.39	-1.6724	ns

* Table 2.1 and Figure 2.2 list full names and locations of sample locations.

† Number of segregating sites

‡ Mean pairwise differences

§ Tajima's (1989) test statistic

[a] Probabilities calculated from confidence limits of *D* given in Tajima (1989)

[1] Tajima's test was applied to the most likely regional division of chimpanzees inferred from other analyses.

[2] Tajima's test was applied to the full HV-I database from chimpanzees in Nigeria and Cameroon.

[3] HV-I sequences of chimpanzees in Upper Guinea were taken from Gagneux (1998).

[4] HV-I sequences from chimpanzees in eastern Africa were taken from Goldberg (1996).

uses an explicit model of evolution for phylogenetic reconstruction. The maximum likelihood approach, therefore, attempts to estimate the *actual* amount of divergence between lineages, rather than the *minimum* amount of change reconstructed during a parsimony analysis (Felsenstein 1981; Felsenstein 1984; Felenstein 1993). Like parsimony analyses, several different trees are examined in a maximum likelihood phylogenetic reconstruction. The optimal tree under the maximum likelihood approach is the tree that is the most likely to have occurred given the observed data and the assumed model of evolution (Felsenstein 1981; Olsen *et al.* 1994; Swofford *et al.* 1996).

Consequently, using an appropriate model of evolution is essential to reconstructing a meaningful phylogenetic tree in the maximum likelihood approach (Felsenstein 1981; Swofford *et al.* 1996). When the correct model of evolution is used, maximum likelihood tends to outperform parsimony in accurately reconstructing phylogenetic relationships, even with very short sequences (Swofford *et al.* 1996). Maximum likelihood is less affected by sampling error, and is robust to violations of the assumptions of sequence evolution models. Additionally, sequence evolution models can be described by just a few parameters under the assumption that the actual substitution processes taking place at different sites have much in common but are not necessarily identical at each site (Hasegawa and Fujiwara 1993; Kuhner and Felsenstein 1994; Huelsenbeck 1995; Swofford *et al.* 1996).

Pairwise distance methods are purely algorithm-based. Pairwise distance methods use corrected distances that account for superimposed changes by estimating the number of unseen events using similar models as those employed in a maximum likelihood analysis (Felsenstein 1984; Swofford *et al.* 1996). There are two steps in distance-based phylogenetic reconstruction. First, a pairwise distance matrix is generated based on a given model of evolution (methods reviewed in: Zharkikh 1994; Swofford *et al.* 1996). After a matrix of distances is generated, a specific tree-building algorithm is used, like neighbor-joining or UPGMA, to generate a single reconstruction from the information contained in the pairwise distance matrix (Sneath and Sokal 1973; Saitou and Nei 1987).

The advantage of pairwise distance methods is that by using different corrected distances, it is possible to take advantage of models simulating a maximum likelihood analysis that are computationally inexpensive. When large data sets are involved, therefore, pairwise distance methods allow more through testing of models of evolutionary change (Swofford *et al.* 1996). However, distance methods are generally considered less desirable approximations of the maximum likelihood approach. When a pairwise distance method is used in a reconstruction, a great deal of information is lost because the entire sequence is reduced to a single-point estimator of divergence (Felsenstein 1984; Huelsenbeck 1995; Swofford *et al.* 1996). Each of these methods was used to analyze the entire HV-I sequence database.

Phylogenetic reconstructions of the HV-I sequence database

Parsimony reconstruction

Theory and methods

Parsimony analyses are divided into two types, exact algorithms and heuristic approaches (Swofford *et al.* 1996). Exact algorithms search through all possible trees to find the most parsimonious solution (i.e., the tree containing the fewest number of steps to explain a given data set). However, using exact algorithms is feasible for data sets containing less than 10-20 taxa because the addition of more samples makes the number of unique bifurcating trees that can be constructed enormous. Consequently, for larger data sets executing a search using an exact algorithm requires a prohibitive amount of computing time (Maddison, Ruvolo, and Swofford 1992; Desalle, Wray, and Abasher 1994; Swofford 1999).

Heuristic approaches, however, sacrifice the guarantee of finding the most parsimonious tree in favor of increased computing speed (Swofford *et al.* 1996). Heuristic tree searches generally operate by hill climbing methods. An initial tree is obtained by one of several methods (i.e., stepwise addition, star decomposition). After an initial tree is constructed, the tree branches are rearranged by one of several branch-swapping methods (i.e., tree bisection-reconnection {TBR}, nearest neighbor interchanges, subtree pruning and regrafting) until the tree reaches a defined optimality criterion (i.e., minimum length) or until the branch-swapping algorithm can no longer improve the tree. However, heuristic searches do not always find the most parsimonious tree. Consequently, multiple replicates of heuristic searches are often used to maximize the amount of "phylogenetic space" examined in a phylogenetic reconstruction (DeSalle, Wray and Abasher 1994; Swofford *et al.* 1996; Swofford 1999).

For a data set of the size (n =254), finding an exact solution is prohibitively time-consuming because the number of unique bifurcating trees that could be reconstructed for a data set this size is very large (Maddison, Ruvolo, and Swofford 1992; Swofford 1999). Until recently, using heuristic approaches to construct trees for large data sets was impractical because it would have required several months to years to execute a statistically meaningful search of "phylogenetic space" (Maddison, Ruvolo, and Swofford 1992; Desalle, Wray, and Abasher 1994; Swofford 1999). However, several new algorithms, computer programs and increased computing power have made it possible to conduct a meaningful search of "phylogenetic space" using heuristic approaches (Goloboff 1999).

The computer program, NONA (Goloboff 1999), can execute heuristic searches of large DNA sequence databases. The algorithms used in the NONA program search for trees that minimize the number of steps needed to reconstruct the most parsimonious tree according to prior weights determined by the user. A number of heuristic search options are available within the program. NONA can execute multiple random addition searches. In addition, a number of branch-swapping algorithms are available including, TBR (Goloboff 1999).

Results and interpretation

The heuristic search options available in the NONA computer program (Goloboff 1999) were used to construct a phylogenetic tree of the entire HV-I sequence database. Figures 3.1 and 3.2 depict a strict consensus of the 40 most parsimonious trees found in a 100 replicate random addition heuristic search using the TBR branch-swapping algorithm. The 20 shortest trees found in each random addition replicate were saved and the branches were swapped by TBR. The 40 most parsimonious trees from the 100 replicates were saved after the completion of the last replicate. The tree is rooted with two humans and two bonobos. Appendix 3 lists sample names for the terminal tree branches.

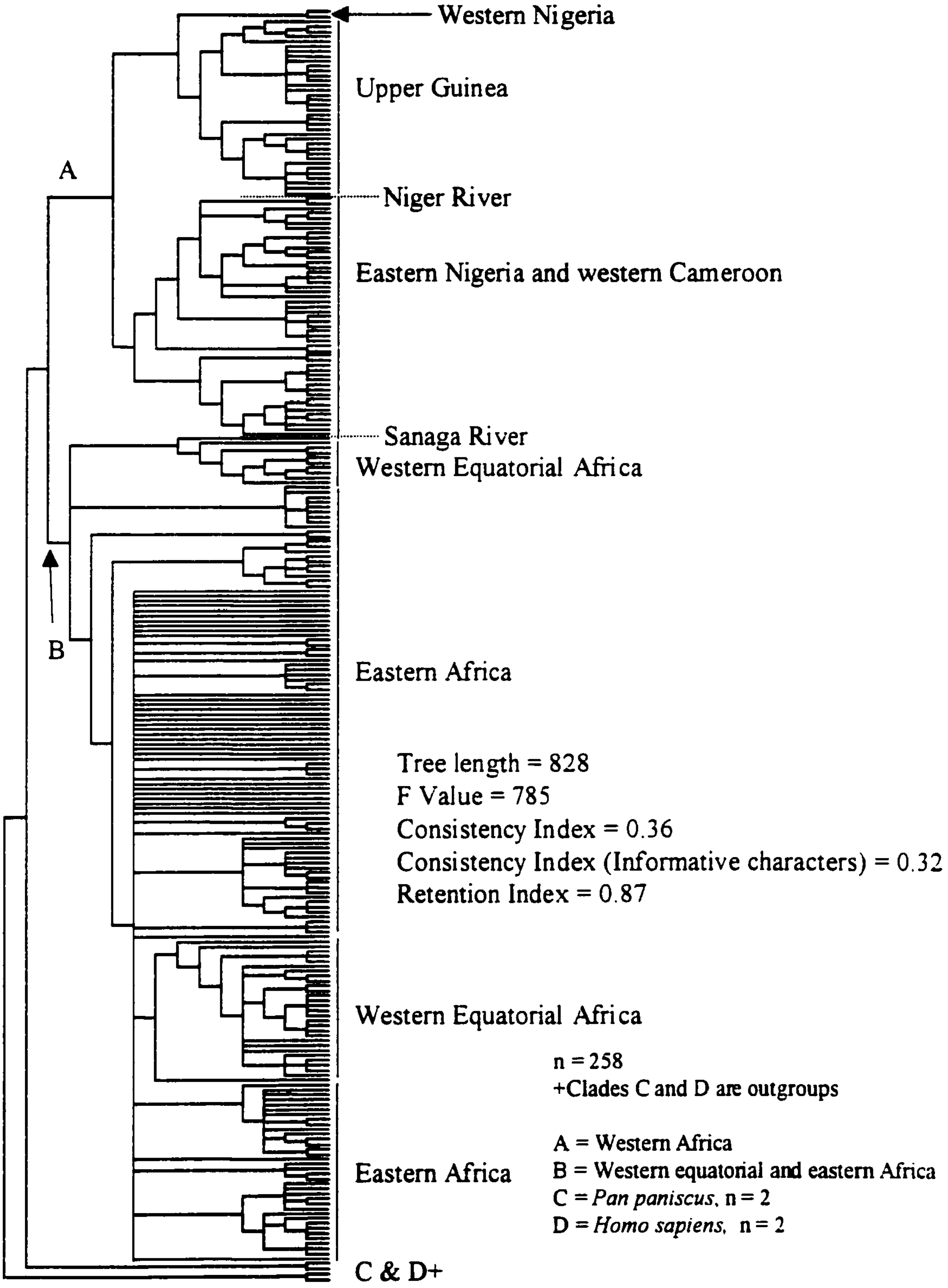


Figure 3.1. Parsimony tree for chimpanzees (*Pan troglodytes*) at the HV-I locus. The phylogenetic reconstruction was performed using the computer program NONA (Goloboff 1999). This tree is a strict consensus of all trees examined in a 100 replicate random addition heuristic search using TBR branch swapping. See text regarding definitions of major geographic divisions. Figure 3.2 and Appendix 3 list sample names and their tree position.

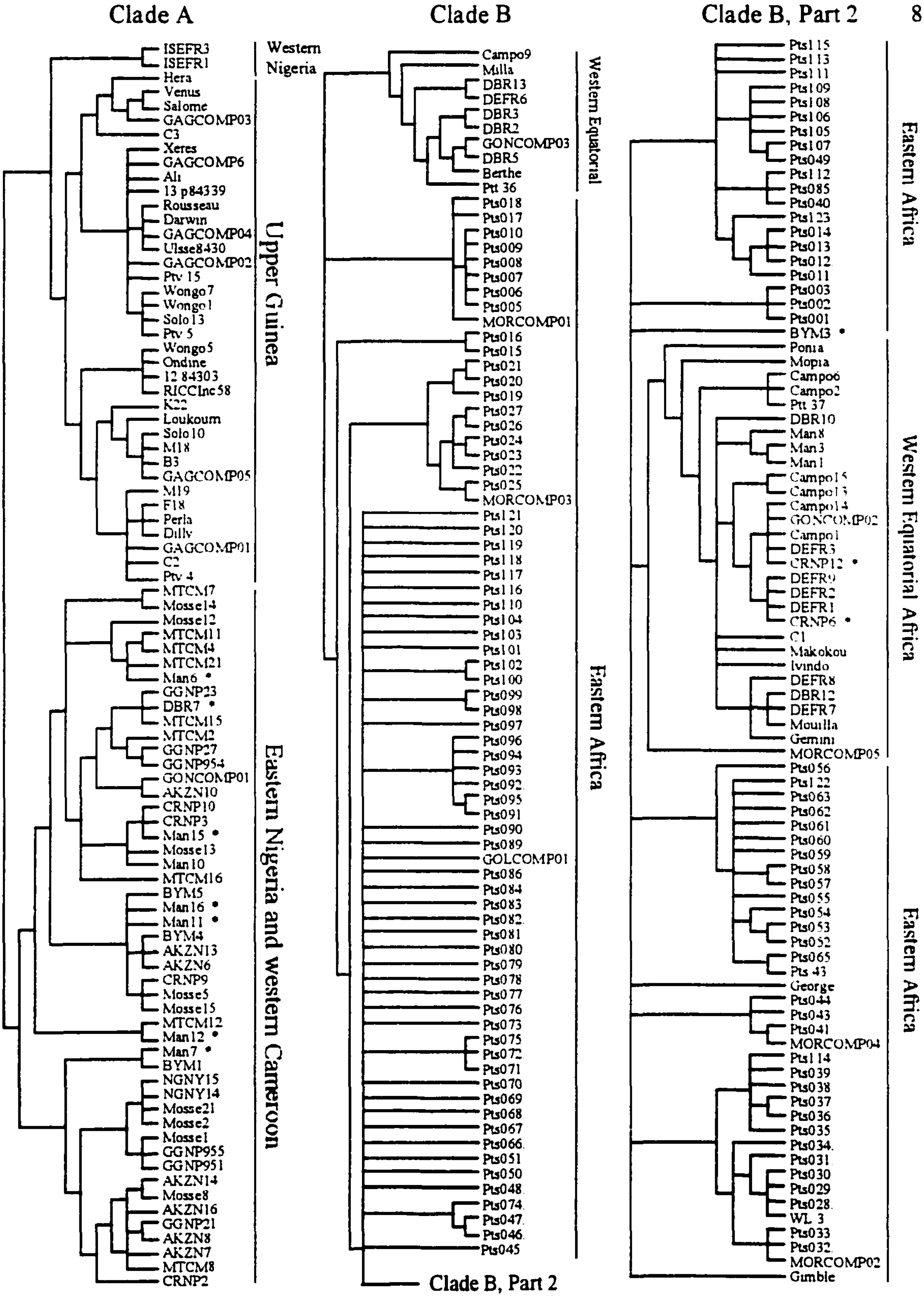


Figure 3.2. Clades A and B of the parsimony tree. Asterisks denote sequences of samples collected outside the geographic area given to the right of the tree. Full tree topology shown in Figure 3.1.

The topology of the phylogenetic reconstruction depicted in Figures 3.1 and 3.2 is not consistent with the generally accepted geographical and taxonomic subdivision of chimpanzees. The parsimony tree suggests that chimpanzees are divided into two large monophyletic clades. The geographic ranges of these putative lineages converge in central Cameroon, near the Sanaga River. Clade A is mostly composed of individual sequences from Upper Guinea, Nigeria and western Cameroon. Clade B is mostly composed of individuals sequences from western equatorial Africa and eastward to the rim of Congo Basin forest expanse in Tanzania. The geographic division of the sequences roughly occurs 10 km north of the Sanaga River at Mosse (map location, 12) and at the Douala-Edea Forest Reserve (Map location, 14), located 20 km south of the Sanaga River. Manb'ra (map location, 13) is located 50 km east the Mbaum River. The Mbaum River is the main tributary of the Sanaga River. Samples from Manb'ra are equally divided between clades A and B.

Within clade A two monophyletic groups are present: one clade composed of individual sequences from Upper Guinea and western Nigeria and one clade composed of individual sequences from eastern Nigeria and western Cameroon to the north bank of the Sanaga River. The top two terminal taxa in Figures 3.1 and 3.2 are both from samples collected in western Nigeria at Ise Forest Reserve (map location, 1). The position of the Ise Forest Reserve samples implies that a phylogeographic break may occur between chimpanzees in western Africa at the Niger River but not at the Dahomey Gap. Further statistical tests to examine the phylogenetic position of these samples were not possible, as the sample size of complete chimpanzee HV-I sequences in Nigeria west of the Niger

River is small (n = 2). One sequence from Dja Biosphere Reserve (DBR7) and seven sequences from Manb'ra are present in the Nigeria-western Cameroon clade.

Clade B primarily is composed of individual HV-I sequences from western equatorial and eastern Africa. However, in contrast to the widely accepted view chimpanzee distribution patterns and phylogenetics, these two regional populations do not cluster into two monophyletic clades. A few individual sequences from the north side of the Sanaga River cluster within this clade. CRNP6 and CRNP12 from Cross River National Park (map locations, 3 and 4) cluster within a western equatorial chimpanzee group. The HV-I sequence of BYM3 from Banyang'mbo (map location, 10) clusters within clade B. Given the surprising location of these samples in the parsimony tree, a second set of fresh DNA extracts from nests hairs, PCR product and HV-I sequences were produced for these samples to reduce the possibility they cluster with sequences from other areas of Africa as a result of contamination.

Maximum likelihood reconstruction

Theory and methods

Maximum likelihood attempts to estimate the *actual* amount of genetic divergence between lineages. In a maximum likelihood analysis, the model of sequence evolution and the algorithm used to evaluate the likelihood of a gene tree are critical in generating an accurate phylogenetic reconstruction. Phylip's Fast DNAml program (Felsenstein 1981; Felenstein 1993; Olsen *et al.* 1994) is often used to reconstruct maximum likelihood trees for very large data sets. Fast DNAml uses a Markov sequence evolution model that assumes two substitution types (transitions vs. transversions) and accommodates unequal base frequencies in a two-parameter model. This model also divides the substitution process into two components: a general substitution rate capable of producing all types of substitutions, and a within-group substitution rate that produces only transitions. This method allows for among-site rate variation to accommodate unequal transition:transversion (TS:TV) ratios across sites (Kishino and Hasegawa 1989; Olsen *et al.* 1994). In the tree building algorithm, the likelihood of an ancestral state is the product of the likelihoods of the state giving rise to the daughter trees, and is a function of the costs of different branch lengths. Maximum likelihood values are then computed for many different tree topologies, and the topology with the highest likelihood is chosen as the most likely candidate for the true tree. The highest likelihood is determined by the combined likelihood for all ancestral states consistent with the tree and branch lengths (Felsenstein 1981; Felsenstein 1984; Goldman 1990; Olsen *et al.* 1994; Swofford *et al.* 1996).

Results and interpretation

Figures 3.3 and 3.4 present a phylogeny of chimpanzee HV-I haplotypes reconstructed using the fast-DNAml program. This tree represents the most likely topology of 218,608 tree topologies examined (*ln* likelihood = -5459.55). The tree was constructed utilizing empirical base frequencies and a 3:1 TS:TV ratio. Several TS:TV ratios were used from 15:1, 10:1, 7:1, 5:1 and 3:1. The resulting tree topology did not change until the 15:1 ratio was used. Appendix 3 lists sample names for the terminal tree branches.

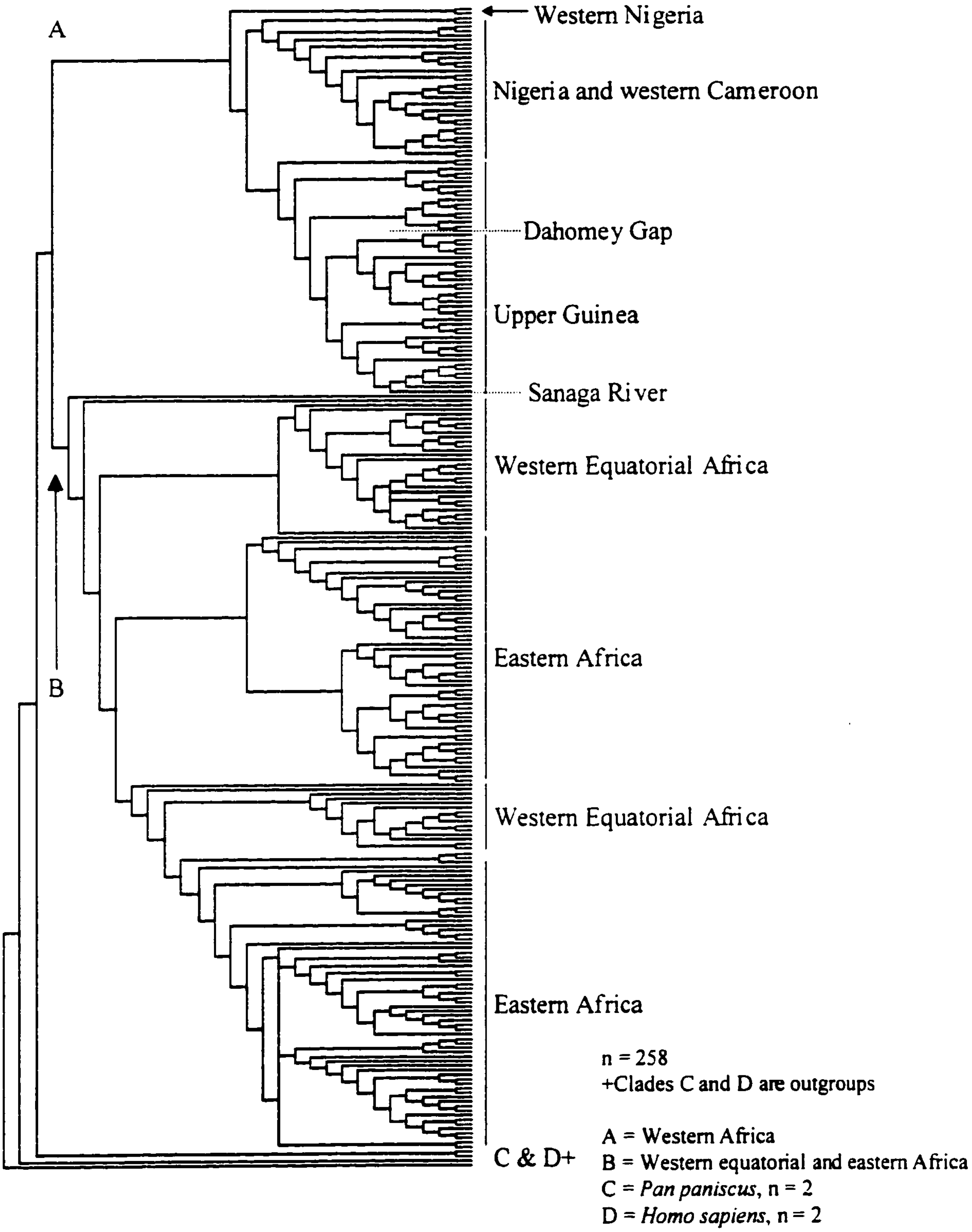


Figure 3.3. Maximum-likelihood tree for chimpanzees at the HV-I locus. The Fast DNAml program (Felsenstein 1991, 1994) was used to construct this tree. See text regarding definitions of major geographic divisions. Figure 3.4 and Appendix 3 list sample names and their tree position.

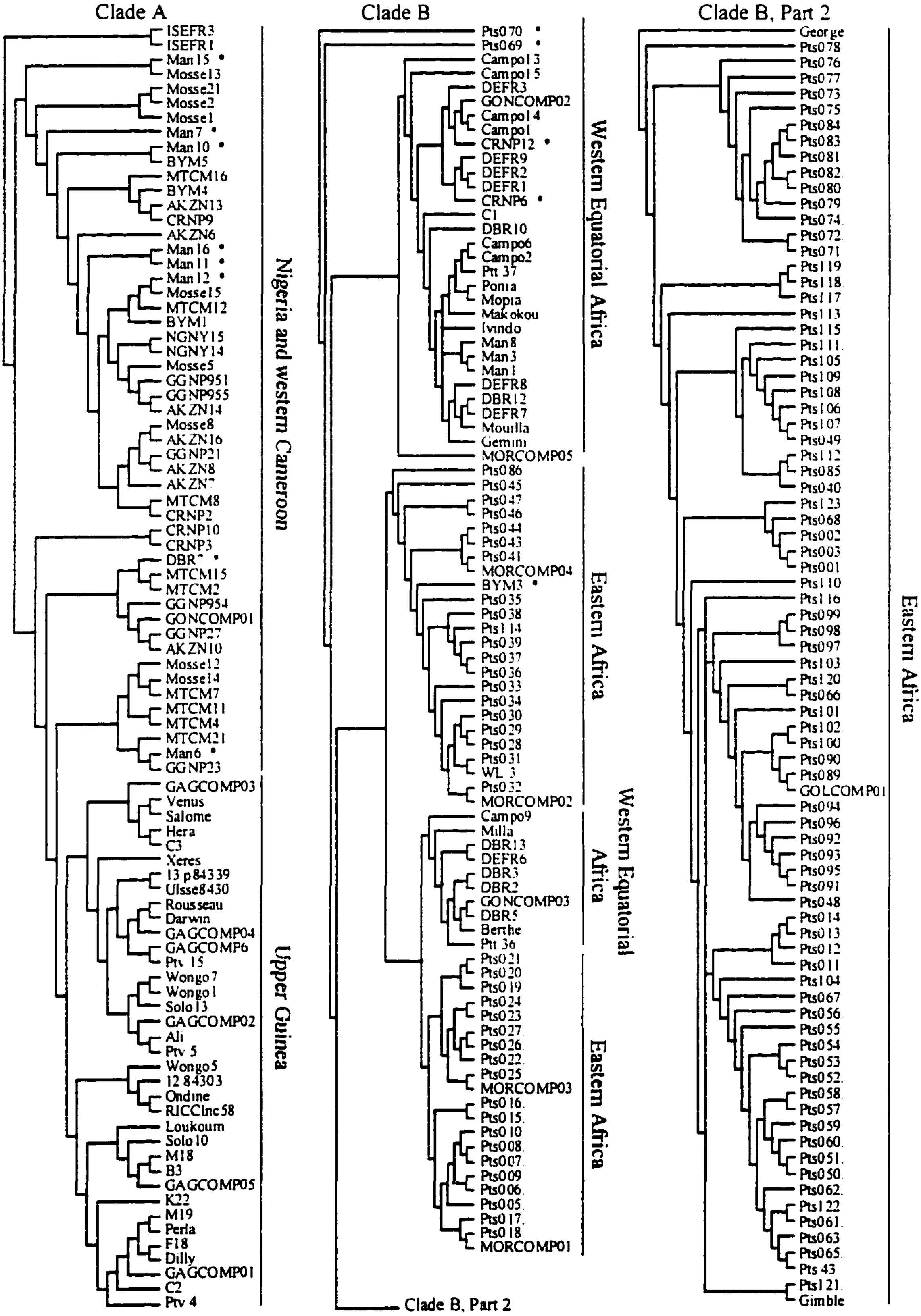


Figure 3.4. Clades A and B of the maximum likelihood tree. Asterisks denote sequences of samples collected outside the geographic area given to the right of the tree. Full tree topology shown in Figure 3.3.

Two large clades are present in the maximum likelihood reconstruction that are geographically divided by the Sanaga River in Cameroon. The sequences present in clades A and B of the maximum likelihood tree are identical to the sequences present in clades A and B of the parsimony tree from Figures 3.1 and 3.2. Consequently, within 100 km of the Sanaga River, sequences cluster into the two major clades defined along the Sanaga but with a few exceptions (CRNP6, CRNP12, BYM3, Campo9 and DBR7). DBR7 clusters within clade A. CRNP6, CRNP12, BYM3 and Campo9 cluster within clade B. Individual sequences from Manb'ra are equally divided between the two clades.

However, the topologies within clades A and B of the maximum likelihood tree are different from those within clades A and B of the parsimony reconstruction. While clade A represents a western African group from Upper Guinea, Nigeria and western Cameroon north of the Sanaga River, the maximum likelihood tree suggests monophyly within this group is not concordant with geography, and therefore, does not support a division of western African chimpanzees into two lineages. Chimpanzees from western Nigeria are basal to all chimpanzees in western Africa. Chimpanzees in Upper Guinea form a monophyletic group within the Nigeria and western Cameroon cluster of Clade A.

Similarly, the branching pattern within clade B is not consistent with geographical proximity or with a geographically nested branching pattern. Sequences from chimpanzees from western equatorial and eastern Africa are present in many of the same small tree branch clusters. Branch lengths also are much shorter within clade B than clade A, suggesting a more recent coalescence of the sequences and perhaps a higher migration rates among these chimpanzees.

Pairwise distance neighbor-joining reconstruction

Theory and methods

Several distance corrections, cluster algorithms and TS:TV ratios were used to evaluate the full data set. However, the Tamura and Nei (1993) distance correction has several properties conducive to more accurately reconstructing HV-I locus gene trees. Similar to other methods, the Tamura and Nei distance correction allows for different TS: TV ratios; but unlike other correction methods it also draws a distinction in transition rates between purines and pyrimidines. Tamura and Nei (1993) showed that at the HV-I locus there are many more pyrimidine (C↔T) substitutions than purine (A↔G) substitutions in hominoid HV-I sequences. Several iterations of their data showed that without this distinction at the HV-I locus, pairwise distance methods are much less robust to violations of sequence evolution models (Tamura and Nei,1993). The Tamura and Nei (1993) distance matrix is defined by:

$$c_1=\frac{2g_Ag_G}{g_R},c_2=\frac{2g_Cg_T}{g_Y},$$

$$c_3=\frac{2g_Ag_Gg_R}{2g_Ag_GgR-g_R^2\hat{P}_1-g_Ag_GQ},c_4\frac{2g_Tg_Cg_Y}{2g_Tg_Cg_Y-g_y^2\hat{P}_2-g_Tg_CQ} \tag{3.2}$$

$$c_5=\frac{2g_Ag_G}{2g_Ag_Gg_R-g_R^2\hat{P}_1-g_Ag_GQ+2g_Tg_Cg_Y-g_y^2\hat{P}_2-g_Tg_CQ}$$

where g_A, g_G, g_C, and g_T represent nucleotide base substitutions; and g_R and g_Y represent purine and pyrimidine substitutions, respectively.

Neighbor-joining is a heuristic method for obtaining a point estimate of a minimum evolution tree (Saitou and Nei 1987), making it impossible to compare optimal and sub-optimal gene trees. However, it is possible to statistically test the reliability of the resulting tree. There are two main resampling options available: nonparametric bootstrap and jackknife resampling procedures (Felsenstein 1985; Hillis 1995; Swofford *et al.* 1996). In a bootstrap resampling procedure, pseudoreplicate data sets are generated by randomly sampling the original character matrix with replacement to create new character matrices of the same size as the original. In bootstrap replicates, some data points will not be included at all in a given bootstrap replicate, while others will be included once, twice or more. Bootstrapping is dependent upon a large number of replicates that incorporate each sample several times. For this data set, unfortunately, well over 100,000 replicates would be necessary to statistically evaluate the position of each sample in a resulting gene tree.

The jackknife procedure, on the other hand, resamples the original data set by dropping k data points at a time and recomputing the estimate from the remaining $n - k$ observations. In the case of nucleotide sequences, a random but specified number of nucleotides are removed from each sample in each replicate (Swofford *et al.* 1996). Thus, the strength of jackknife resampling is that the original data set is maintained but the number of bases analyzed varies in each replicate. This property lowers the number of replicates necessary to reconstruct a phylogenetically meaningful tree.

Results and interpretation

Figures 3.5 and 3.6 present a phylogeny of chimpanzee HV-I sequences reconstructed using a matrix of pairwise distances generated using the Tamura and Nei (1993) distance correction, and a TS:TV ratio of 3:1. A neighbor-joining algorithm was

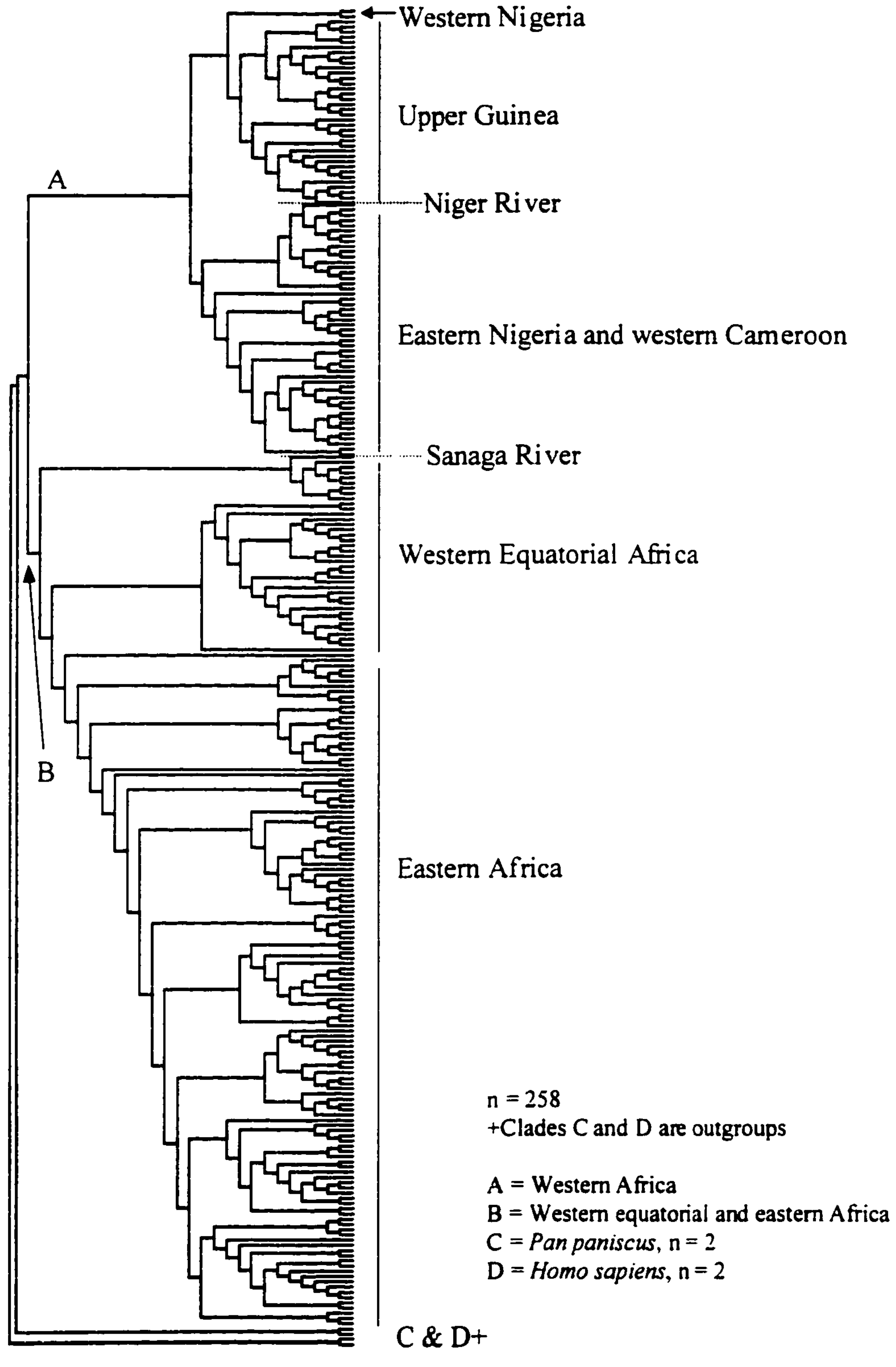


Figure 3.5. Jackknife neighbor-joining tree for chimpanzees at the HV-I locus (10,0000 replicates). The phylogenetic reconstruction was performed using the Tamura and Nei (1993) distance correction. See text regarding definitions of major geographic divisions. Figure 3.6 and Appendix 3 list sample names and their tree position.

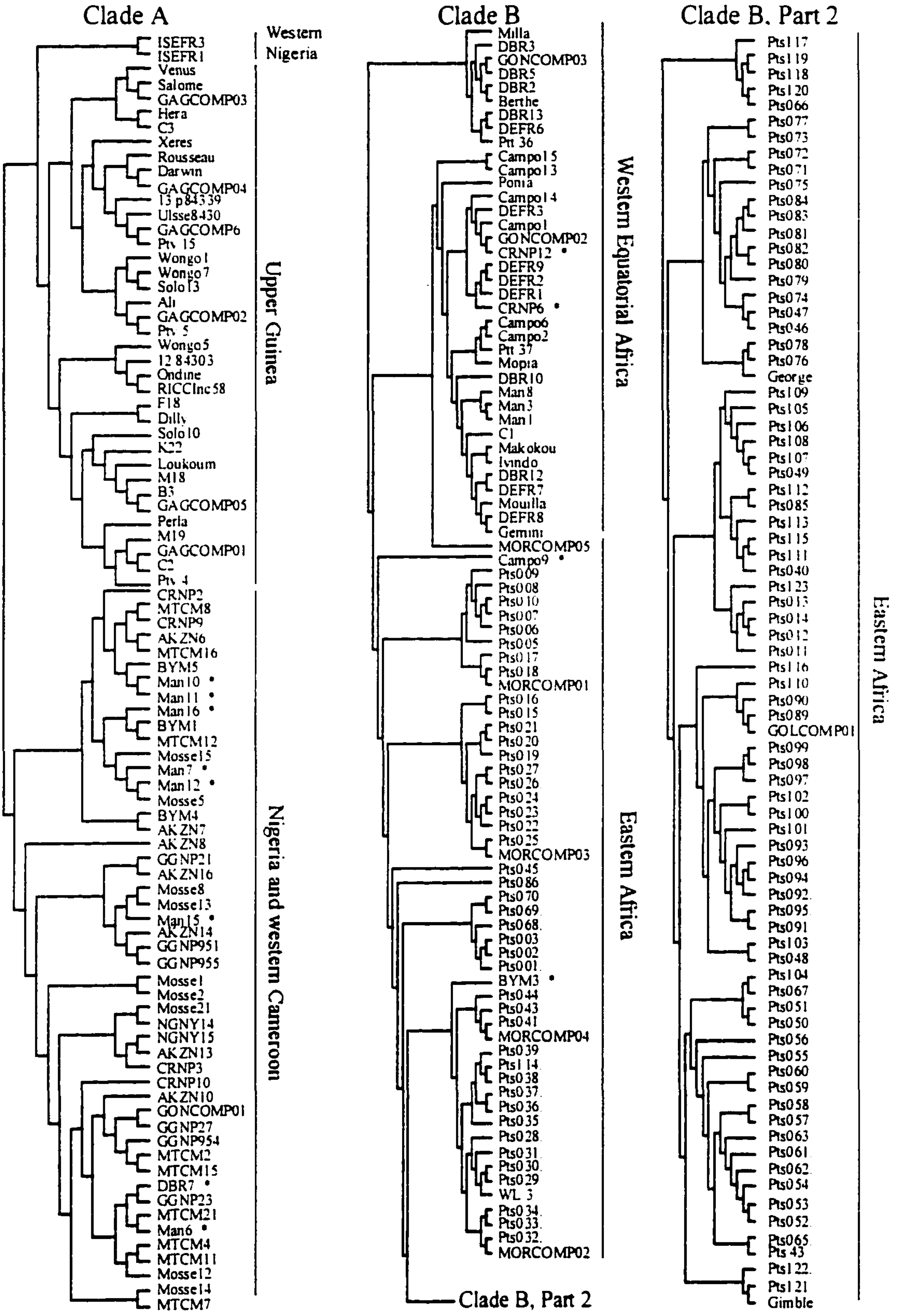


Figure 3.6. Clades A and B of the jackknife neighbor-joining tree. Asterisks denote sequences of samples collected outside the geographic area given to the right of the tree. Full tree topology shown in Figure 3.5.

used to generate the gene tree in PAUP* (Swofford 1999). 10,000 jackknife replicates with 50% deletion in each replicate were used to reconstruct the tree with confidence limits. Appendix 3 lists sample names for the terminal tree branches.

The topology of the jackknife neighbor-joining tree is almost identical to the parsimony reconstruction from Figure 3.1. Branch support for the clades within the neighbor-joining tree was high ranging from 85-100%. These results provide further support that chimpanzees are divided into two large monophyletic clades. Similar to the parsimony and maximum likelihood reconstructions, clade A is mostly composed of individual sequences from Upper Guinea, Nigeria and western Cameroon. Clade B is mostly composed of individuals sequences from western equatorial and eastern Africa. Moreover, this phylogenetic reconstruction also implies that the geographic division of the sequences occurs in the vicinity of the Sanaga River between Mosse (map location, 12), and the Douala-Edea Forest Reserve (Map location, 14). Samples from Manb'ra (map location, 13) are equally divided between clades A and B.

Within clade A two monophyletic groups are present: one clade composed of individual sequences from Upper Guinea and western Nigeria and one clade composed of individual sequences eastern Nigeria and western Cameroon. Similar to the parsimony reconstruction, samples from western Nigeria (ISEFR1 and ISEFR3) form part of group composed of chimpanzees in Upper Guinea. These findings provide further support that a phylogeographic break between chimpanzees in western Africa occurs at the Niger River, but not at the Dahomey Gap. Clade B is primarily composed of individual HV-I sequences from western equatorial and eastern Africa. Similar to the parsimony and maximum likelihood trees, samples from chimpanzees in each regional population do not form two monophyletic clades within clade B. However, the branching patterns within clade B are nested geographically.

Although the jackknife neighbor-joining tree suggests a major phylogeographic division of chimpanzees in the vicinity of the Sanaga River, there are a few exceptions (CRNP6, CRNP12, BYM3, Campo9 and DBR7). DBR7 clusters within clade A. CRNP6, CRNP12, BYM3 and Campo9 cluster within clade B. CRNP6 and CRNP12 cluster with samples from chimpanzees in western equatorial Africa. BYM3 and Campo9 cluster with chimpanzees from eastern Africa. These findings imply that that although the Sanaga River may influence chimpanzee distribution patterns, it may not completely inhibit gene flow between chimpanzee populations.

Character analysis of the phylogenetic reconstructions

The computer program MacClade (Maddison and Maddison 1996) was used to examine the nucleotide substitutions that divided the clades in the phylogenetic reconstructions in shown Figures 3.1-3.6. Character state changes were traced along the branches of each gene tree. Using this method, five fixed substitutions (1 transition and 4 transversions) support the branch dividing clades A and B; and several substitutions are nearly fixed in each clade. Within clade A, the branch separating the clade composed of chimpanzees in Upper Guinea and those in eastern Nigeria to western Cameroon is supported by 4 fixed transitions in both the parsimony and distance-based phylogenetic reconstructions. Several more substitutions are nearly fixed within each group, including one transversion that is fixed in all but two sequences (NGNY14 and NGNY15). The Ise Forest Reserve samples (map location, 1) were excluded from these counts because the maximum likelihood reconstruction placed them as basal members of clade A. At the fixed nucleotide substitutions, these samples share three substitutions with chimpanzees from Upper Guinea (2 transitions and 1 transversion) and two transition substitutions with chimpanzees in eastern Nigeria and western Cameroon. The Ise samples also contain four fixed substitutions that are different all other chimpanzees in western Africa (3 transitions and 1 transversion). No fixed substitutions were detected that divide HV-I sequences of chimpanzees in western equatorial and eastern Africa. However, two transition substitutions are nearly fixed differences between these regional populations.

Molecular evolution at the HV-I locus

Theory and methods

Transitional and transversional substitutions tend to occur at unequal frequencies along DNA sequences. In most cases, transitional substitutions tend to accumulate more rapidly than transversional substitutions (Li and Graur 1991; Ruvolo *et al.* 1993). Because transitions accumulate much more rapidly than transversions at a given locus, it can be difficult to have a sufficient number of transversions to infer a branching pattern without having the phylogenetic signal randomized by transitions (Swofford *et al.* 1996). This problem is especially pronounced if the time period in which phylogenetically informative mutations might arise is very narrow, as is generally the case among closely related lineages (Simon 1991; Ruvolo *et al.* 1993).

As the phylogenetic distance between sequences increases the TS:TV ratio tends to decrease (Swofford *et al.* 1996). For example, analyses of HV-I sequences of human family members suggest that the inferred mutation rate and TS:TV ratio is twenty times the estimated mutation rate and TS:TV ratio inferred from phylogenetic studies (Howell, Kubacka, and Mackey 1996; Parsons *et al.* 1997). Moreover, Meyer, Weiss and von Haeseler (1999) have shown that at the human HV-I locus the TS:TV ratio decreases with increasing sample size (and presumably with increased phylogenetic distance between sequences). In a rapidly evolving region like the HV-I locus, a decrease in the TS:TV ratio increases the likelihood that multiple substitutions may have occurred at the same nucleotide position along the sequence (Li and Graur 1991; Swofford *et al.* 1996). As a result, the least-biased estimates of the TS:TV ratio are obtained as the phylogenetic distance between sequences reaches zero (Ruvolo *et al.* 1993; Yang and Yoder 1999). In response to these problems, models of sequence evolution use a TS:TV ratio to reduce the "noise" transitional substitutions can introduce into a phylogenetic analysis (Ruvolo *et al.* 1993; Swofford *et al.* 1996; Ina 1998; Meyer, Weiss, and von Haesler 1999). When TS:TV ratios are designated in a phylogenetic reconstruction a step matrix is created so that different categories of change are weighted differently. For example, if a TS:TV ratio of 10:1 is used in a phylogenetic reconstruction, each transitional change will alter the tree length by one step, while each transversional change will add ten steps to the total tree length. Thus, changing this ratio can alter many aspects of a phylogenetic reconstruction including, minimum and maximum tree lengths and decay index support values (Bremer 1994; Swofford *et al.* 1996).

The Arlequin computer program (Schneider *et al.* 1997) can implement a procedure to estimate empirical TS:TV ratios by counting the number of transitional and transversional substitutions in an alignment of multiple DNA sequences from a reference DNA sequence. This procedure was used to calculate TS:TV ratios for the entire HV-I sequence data base. These calculations indicated an empirical pan-African TS:TV ratio of 3:1. Similarly, the overall TS:TV ratio for chimpanzee HV-I sequences from Nigeria and Cameroon was estimated at 3:1. This TS:TV ratio contrasts with the TS:TV ratios given in other studies. Morin *et al.* (1994) used a TS:TV ratio of 15:1 for their pan-African sample; while among chimpanzees in Upper Guinea the TS:TV ratio was estimated at 9.2:1 (Gagneux, 1998). Among eastern African chimpanzees, the TS:TV ratio was estimated at 10:1 (Goldberg, 1996).

Given the differences in the TS:TV ratio calculated from this study and calculated from other studies, patterns of nucleotide substitution inferred by the phylogenetic reconstructions were checked for consistency against the empirical 3:1 TS:TV ratio. Character state changes were traced along each gene tree in 10bp intervals using the computer program MacClade (Maddison and Maddison 1996). In each case, the TS:TV ratios inferred from the gene trees were similar to the ratio detected by counting nucleotide variation along the sequence.

Results and interpretation

Representative results from the maximum likelihood reconstruction are shown in Figures 3.7 and 3.8. The top panel of Figure 3.7 shows the pattern of nucleotide variation for the full HV-I sequence database. The bottom panel of Figure 3.7 shows the pattern of nucleotide variation from an analogous region of the human HV-I sequence with an inferred 17:1 TS:TV ratio (Vigilant *et al.* 1991; Goldberg 1996). The patterns of nucleotide variation detected in these two samples are different. Several more transitional and transversional substitutions have accumulated along the chimpanzee

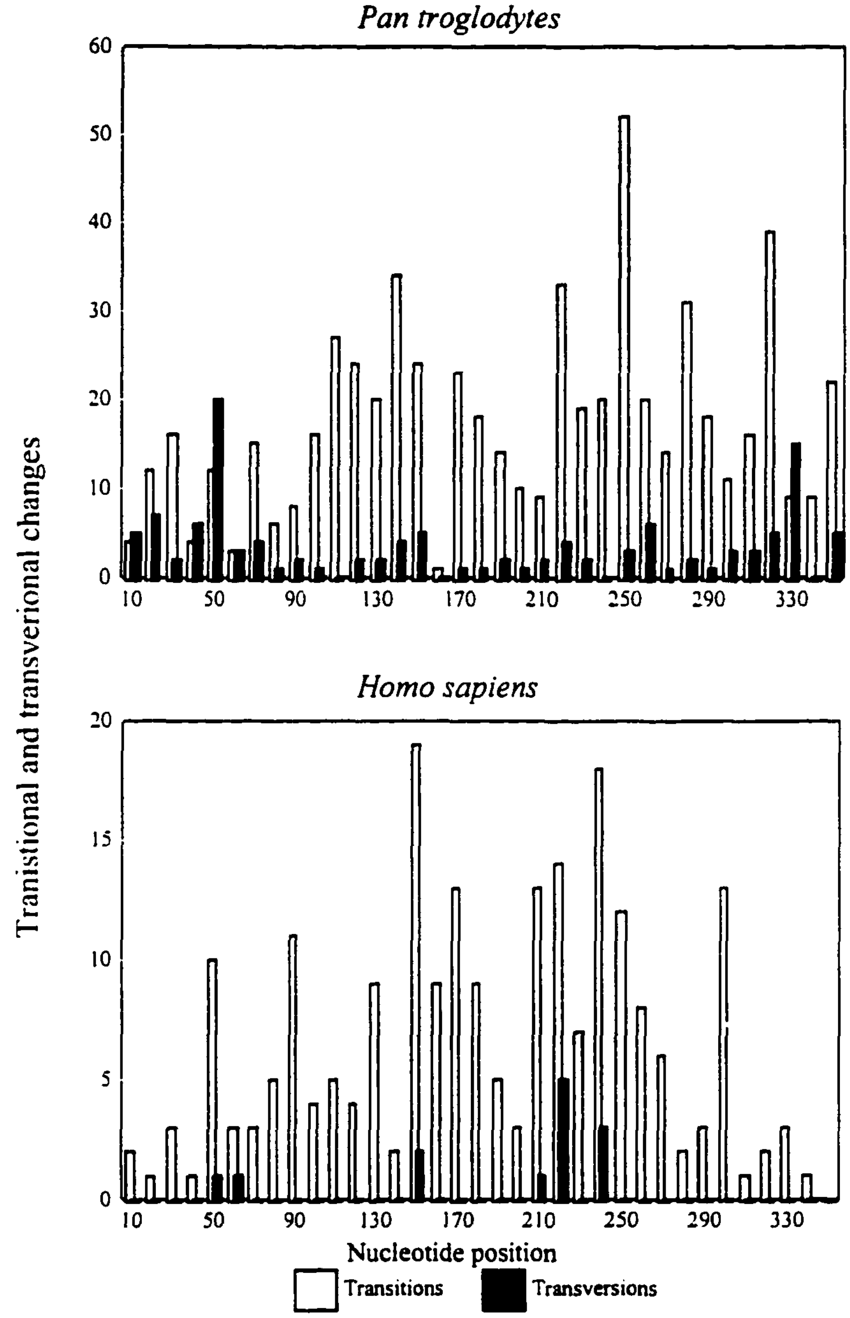


Figure 3.7. Nucleotide substitutions along the HV-I locus in chimpanzees and humans. The minimum numbers of transitions and transversions were calculated in 10-nucleotide intervals using the program MacClade (Maddison and Maddison 1996). In the chimpanzee HV-I data set (n = 254), changes were traced along the branches of the maximum likelihood tree shown in 3.3. Results were similar when nucleotide changes were traced along other trees and by counts of nucleotide variation along the sequence. Human data (n = 135) were taken from Vigilant *et al.* (1991); inferred TS and TV changes were taken from Goldberg (1996).

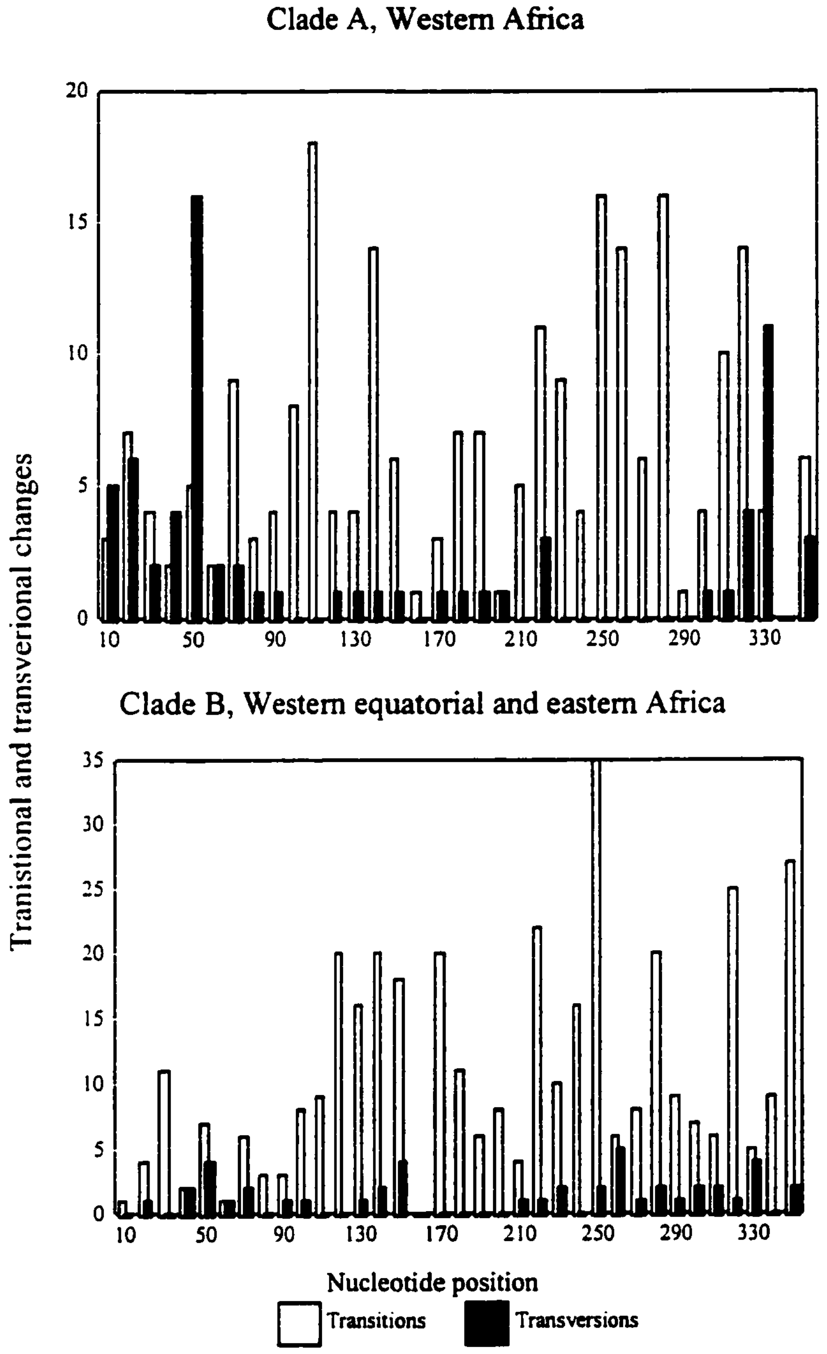


Figure 3.8. Nucleotide substitutions along the HV-I locus in the western and central African chimpanzee lineages. The minimum numbers of transitions and transversions were calculated in 10-nucleotide intervals using the program MacClade (Maddison and Maddison 1996). Changes were traced along the branches of clade A (n = 87) and clade B (n = 167) in the maximum likelihood tree shown in Figure 3.3. Results were similar when nucleotide changes were traced along other trees and by counts of nucleotide variation along the sequence.

HV-I sequence than along the human HV-I sequence. However, Goldberg (1996) showed that the pattern of nucleotide variation at the HV-I locus among chimpanzees in eastern Africa is not different from the pattern observed at the human HV-I locus. These differences in nucleotide variation patterns imply, therefore, that the branches separating clades A and B might be deeper than those dividing human populations.

However, within clades A and B the relationships between chimpanzees may be more recent. Figure 3.8 compares the patterns of nucleotide substitution between clades A and B of the maximum likelihood tree shown in Figure 3.3. The pattern of nucleotide variation is different between the clades. Transversional substitutions account for many more of the genetic differences detected in clade A than in clade B. Because transversions accumulate much more slowly than transitions, the chimpanzees represented in clade A might share a more distant relationship with each other than those in clade B.

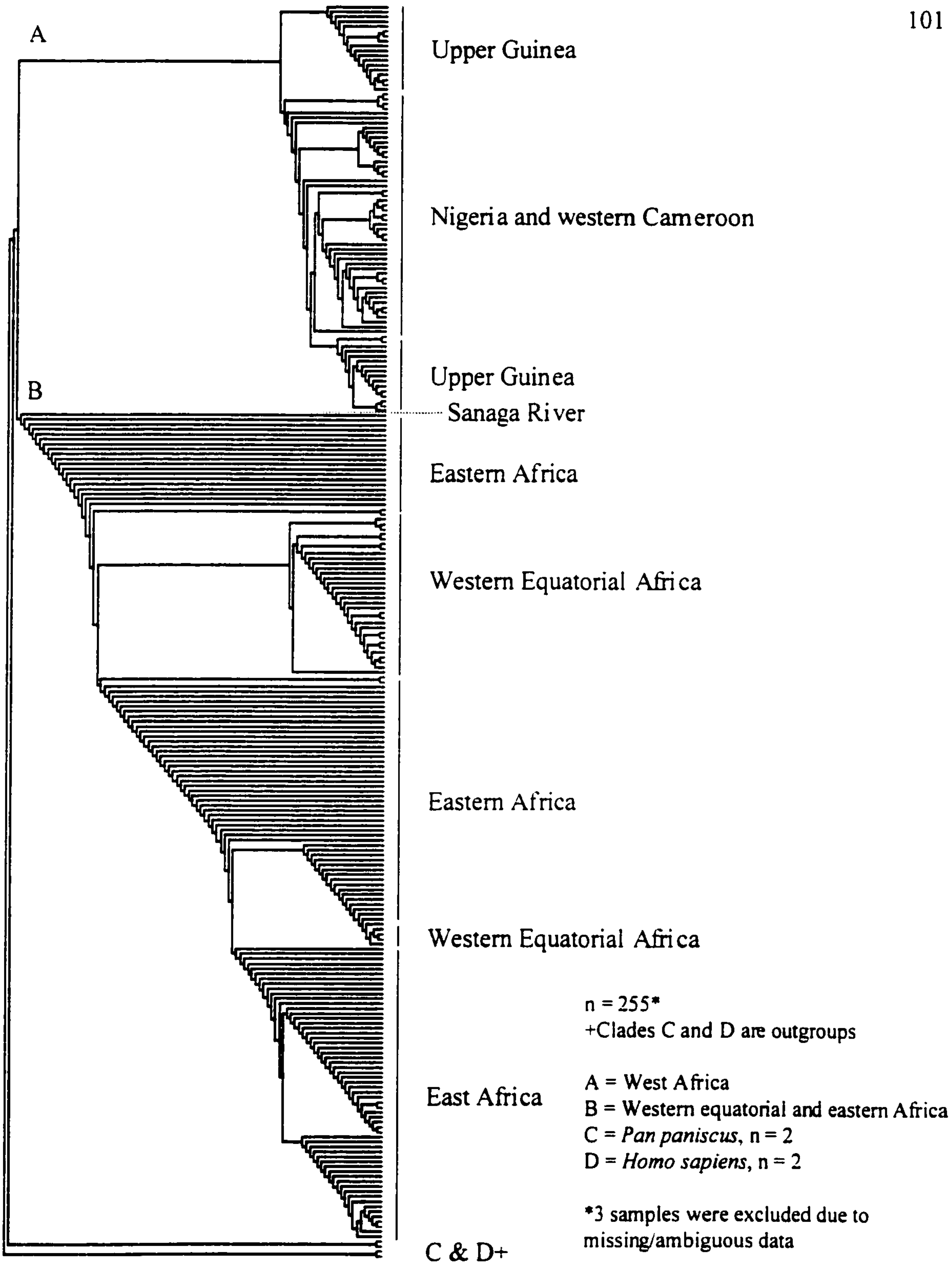
Phylogenetic reconstructions using different substitution classes

Theory and methods

There are two possible explanations of why the patterns of molecular evolution at the HV-I locus differs between chimpanzees and humans and between putative chimpanzee lineages. First, the sequences generated during this study might be faulty. This explanation is unlikely because quadruplicate reads confirmed each base pair change. Moreover, Appendix 2 demonstrates that the sequences from this study align well with sequences from other studies. Alternatively, the inferred TS:TV ratio may be an underestimate of the actual TS:TV ratio at the HV-I locus because the large sequence database represents two very ancient lineages. One way to test this hypothesis is to reconstruct a phylogenetic tree using only transversions (Goldstein and Pollock 1994; Tajima and Takazaki 1994; Swofford *et al.* 1996). If the putative older clades are still present in the reconstructed tree, then those lineages are probably very ancient. In that case, transitional substitutions dividing the ancient clades probably should be ignored because those sites may have been "saturated" by multiple substitutions. However, more recent subdivisions within those ancient clades may only be detected by examining more rapidly-accumulating transitional substitutions that are less likely to have undergone multiple substitutions between younger lineages (Goldstein and Pollock 1994; Swofford *et al.* 1996).

Results and interpretation

Figures 3.9 and 3.10 depict a neighbor-joining tree of the entire data set constructed from uncorrected pairwise distances using only transversions. Appendix 3 lists sample names for the terminal tree branches. Like the other phylogenetic reconstructions of the HV-I locus, the neighbor-joining transversion tree defines two major clades of chimpanzees that are geographically separated by the Sanaga River in central Cameroon. However, there is little geographic resolution of the samples within clades A and B. A neighbor-joining tree also was reconstructed using a pairwise distance matrix based on transitional substitutions for the entire data set. The resulting topology was nearly identical to the parsimony and neighbor-joining jackknife trees presented in Figures 3.1 and 3.5. These findings suggest a relatively ancient relationship between



n = 255*
+Clades C and D are outgroups

A = West Africa
B = Western equatorial and eastern Africa
C = *Pan paniscus*, n = 2
D = *Homo sapiens*, n = 2

*3 samples were excluded due to missing/ambiguous data

Figure 3.9. Neighbor-joining tree using a pairwise distance matrix constructed from transversion substitutions. See text regarding definitions of major geographic divisions. Figure 3.10 and Appendix 3 list sample names and their tree position.

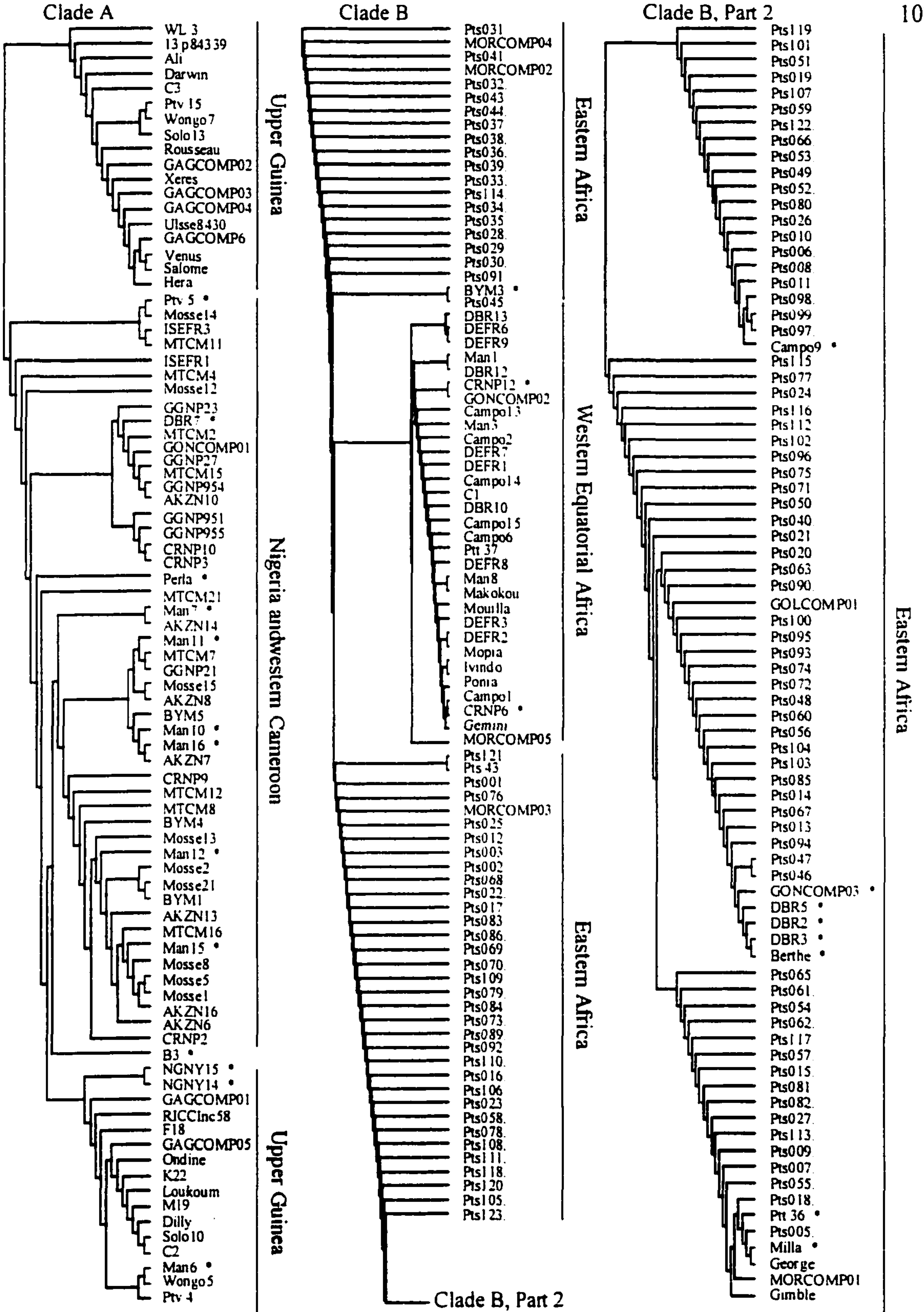


Figure 3.10. Clades A and B of the neighbor-joining transversion tree. Asterisks denote sequences of samples collected outside the geographic area given to the right of the tree. Full tree topology shown in Figure 3.9.

clades A and B and a more recent separation of chimpanzees within clades A and B. Consequently, transitional substitutions may be more useful in differentiating the relationships between populations within clades A and B.

Relative divergence dates

Theory and methods

There are several methods to estimate divergence times and coalescence dates (e.g., Templeton 1993; Vigilant *et al.* 1991). However, calculating these estimates requires, (1) an assumption of rate constancy (i.e., a molecular clock), (2) accurate estimates of inferred amounts of genetic change and (3) an independent calibration point that is usually inferred from evidence found in the fossil record (Swofford *et al.* 1996). Ruvolo *et al.* (1993) have proposed a method of calculating divergence dates that does not require an independent calibration point. They propose estimating *relative* rather than *absolute* divergence times. The relative divergence date method estimates a relative time for the origin of a given clade with reference to an outgroup with an estimated absolute divergence date given in years (Ruvolo *et al.* 1993).

In this method, the number of substitutions between samples of a clade is divided by the number of substitutions between samples in an ancestral clade. The relative divergence date is expressed proportionally to the inferred divergence date of the ancestral clade. All substitution rates are corrected for multiple substitutions by a specified distance correction method such as, the Tamura and Nei (1993) model. This method results in estimating the *upper limit* of the age of a clade, or in genetic terms, an inferred ancestral haplotype. Ruvolo *et al.* (1993) have applied this approach to human mitochondrial DNA sequence data. They proposed a last common human mitochondrial ancestor 298kya based on a putative split between human and chimpanzee species 6mya (Ruvolo *et al.* 1993).

Results and interpretation

The Ruvolo *et al.* (1993) relative divergence date method was applied to the chimpanzee HV-I sequence database. The results are summarized in Table 3.2. All divergence dates are given proportionally to the inferred the human HV-I mitochondrial ancestor 298kya (Ruvolo *et al.* 1993). Confidence limits were calculated using the maximum and minimum number of substitutions for each putative lineage.

Several human HV-I sequences were obtained from Genbank. Using the Tamura and Nei (1993) distance correction and TS:TV ratio of 10:1, these sequences differed by an average of 23 substitutions (21 transitions and 2 transversions). In contrast, genetic diversity was much higher among chimpanzees. Using the Tamura and Nei (1993) distance correction and the empirical 3:1 TS:TV ratio calculated for the entire HV-I database yielded an average of 69 ± 30 substitutions between HV-I sequences (52 transitions and 17 transversions). The approximate origin of the chimpanzee HV-I mitochondrial is therefore, 69/23, or 300% of the human estimate. Consequently, an inferred 298kya last common human mitochondrial ancestor implies a chimpanzee last common mitochondrial ancestor 894kya ± 406ky.

This method was also applied to the major chimpanzee clades inferred from the phylogenetic reconstructions from Figures 3.1-3.6 using distances corrected by the Tamura and Nei (1993) model and empirical TS:TV ratios estimated by the method

Table 3.2 Relative divergence dates for humans and chimpanzees

HV-I Sequences	Substitutions*	Transitions	Transverions	TS:TV†	% of Human estimate‡	Divergence Date (kya)§
Homo sapiens	23	21	2	10:1		298
Pan troglodytes	69 ± 30	52	17	3:1	300	894 ± 406
Upper Guinea	16 ± 7	14	2	9:1	70	207 ± 91
Eastern Nigeria and Western Cameroon	18 ± 8	13	5	3:1	78	233 ± 104
Western Equatorial	24 ± 11	18	6	3:1	104	311 ± 142
Eastern	9 ± 4	7	1	6:1	39	117 ± 51

*Average substitution rate including transitions and tranversions

†Empircal TS:TV ratio calculated for each putative lineage using the method described on P. 95

‡This percentage was calculated by dividing the average number of substitutions by the number of substitutions calculated from the human HV-I sequences

§All divergence dates for chimpanzees are expressed proportionally to the human mitochondrial ancestor 298kya inferred by Ruvolo *et al.* (1993).

described on p. 95. Relative divergence dates suggest that chimpanzees in western equatorial Africa possess the most ancient HV-I sequences. The relative divergence date inferred from the HV-I sequences implies a mitochondrial ancestor 311kya ± 142ky for chimpanzees in western equatorial Africa. Chimpanzees in Upper Guinea and those eastern Nigeria and western Cameroon may have shared more recent mitochondrial ancestors. The relative divergence date estimates of the HV-I sequences imply an origin of Upper Guinea chimpanzees 207kya ± 91ky, while those eastern Nigeria and western Cameroon may have shared a mitochondrial ancestor 233kya ± 104ky. Chimpanzees in eastern Africa may have shared the most recent mitochondrial ancestor. The relative divergence date implies an origin of eastern African chimpanzees 117kya ± 51ky.

Conclusions

The geographical patterning of genetic diversity at the HV-I locus is quite different from the commonly proposed separation of three chimpanzee lineages by the Niger and Ubangi Rivers. The overall patterns of the phylogenetic reconstructions presented in this chapter suggest that there are two deeply divergent chimpanzee lineages: (1) a western African lineage represented by samples from Upper Guinea, Nigeria and western Cameroon, and (2) a central African lineage represented by samples from western equatorial Africa to the eastern edge of the species' range. The differences between these two lineages is well-supported in each phylogenetic reconstruction by four slowly-evolving transversion substitutions.

The phylogenetic reconstructions strongly suggest that the ranges of the western and central African lineages are mostly, but not entirely, delimited by the Sanaga River in central Cameroon. Geographically, this genetic division is so sharply defined that samples collected just north of the Sanaga River at Mosse (map location, 12) and just south of the Sanaga River at the Douala-Edea Forest Reserve (map location, 14) are part of the western and central African lineages, respectively. However, there is some evidence of gene flow between these lineages. HV-I sequences belonging to both lineages were detected in the samples collected at Manb'ra (map location, 13), and haplotypes from both lineages also occur in the Cross River National Park (map locations, 3 and 4) and in the Dja Biosphere Reserve in southern Cameroon (map location, 16). These findings imply that the Sanaga River does not completely deter gene flow between the western and central African lineages.

The phylogenetic reconstructions suggest that the geographical relationships between HV-I sequences within the western and central African clades are less clearly resolved and that chimpanzees within these lineages may share a recent relationship. The jackknife neighbor-joining tree most clearly resolves the genetic differences between western equatorial and eastern African chimpanzees, but reveals only a nested geographic distribution of HV-I sequences between these two regional populations. The other phylogenetic reconstructions also suggest that chimpanzees in western equatorial and eastern Africa are not divided into two monophyletic clades; there is a remarkable lack of geographical resolution in the branch topology of those trees. Moreover, there are no fixed nucleotide differences at the HV-I locus between chimpanzees in western equatorial and eastern Africa. These findings imply a very recent relationship between chimpanzees in western equatorial and eastern Africa.

In contrast, chimpanzees in Upper Guinea and those in Nigeria and western Cameroon probably have had a more complex and ancient relationship. Gonder *et al.* (1997) proposed two putative chimpanzee lineages in western Africa: one limited to Upper Guinea and one limited to Nigeria and adjacent parts of Cameroon. Both the parsimony and neighbor-joining trees support that hypothesis, and imply that chimpanzees in Upper Guinea and those in eastern Nigeria and western Cameroon are part of two monophyletic clades within the western African lineage. However, the maximum likelihood reconstruction and the neighbor-joining tree constructed using a transversion pairwise distance matrix fail to resolve the genetic distinctions between these two putative lineages. The lack of agreement between these phylogenetic reconstructions makes the genetic differences between these two clades ambiguous. Nevertheless, there are more genetic differences between chimpanzees in Upper Guinea and those in eastern Nigeria and western Cameroon than between those in western equatorial and eastern Africa. For example, five fixed substitutions in the parsimony and neighbor-joining jackknife trees separate these two putative lineages, despite the fact that the maximum likelihood and transversion neighbor-joining trees reveal little genetic structure that is related to the geography of western Africa.

The relative divergence date estimates for all the HV-I sequences imply an ancient chimpanzee mitochondrial ancestor (894kya ± 406ky). Chimpanzees in western equatorial Africa possess the most ancient HV-I haplotypes of all modern chimpanzee populations. HV-I haplotypes of chimpanzees from this region may have shared a mitochondrial ancestor 311kya. HV-I sequences of chimpanzees in Upper Guinea, Nigeria and western Cameroon imply a more recent origin of the western African lineage.

HV-I sequences of chimpanzees in Nigeria and western Cameroon imply a divergence of these populations from other chimpanzees roughly 233kya, while those in Upper Guinea may have shared a mitochondrial ancestor about 207kya. Chimpanzees in eastern Africa represent the most youthful group. The relative divergence date estimates implies an origin of eastern African chimpanzees roughly 117kya.

These observations imply that biogeographic boundaries like the Sanaga River, the Niger River and the Dahomey Gap may have had a long term influence on chimpanzee distribution patterns and lineage relationships. The influence of the Sanaga River has probably been significant in recent chimpanzee evolution. However, the importance of the Niger River and the Dahomey Gap in influencing chimpanzee distribution patterns is unclear because the phylogenetic position of chimpanzees in western Nigeria is ambiguous.

Unfortunately, only two complete sequences of chimpanzees from western Nigeria were available for analysis. These two sequences cluster either as part of a group represented by chimpanzees in Upper Guinea or as basal members of the entire western African lineage. Chimpanzees from western Nigeria may be very different from all other western African chimpanzees or part of the group from Upper Guinea. However, the Ise chimpanzee population is bordered on all sides by agricultural lands (Agbelusi 1994; pers. obs.). This small population may have been isolated for a considerable time, making their HV-I sequences appear more divergent than they would if they were not isolated from other populations.

Yet, the genetic differences between chimpanzees in western Nigeria and other chimpanzees in western Africa are probably not entirely attributable to human-induced population isolation. Palynological records suggest that widespread agriculture began in western Africa only about 3kya (Hamilton 1992), and intensive commercial agriculture was only introduced to western Africa during the 1800s (Oates 1999). Consequently, the *maximum* time these chimpanzees could have been influenced by agricultural practices is only on the order of tens of generations, and probably much less. The sample size of chimpanzees from western Nigeria was too small to permit rigorous statistical testing of that hypothesis. Consequently, the full extent of the influence of the Dahomey Gap and of the Niger River on recent chimpanzee evolution cannot be fully evaluated at this time.

CHAPTER 4

Population Genetics of the mtDNA Hyper-Variable Region-I of Chimpanzees in Western Africa

Introduction

The phylogenetic analyses presented in Chapter 3 suggest that there are considerable restrictions to gene flow between chimpanzee populations, particularly near the Sanaga River in Cameroon. However, it is not possible to quantify the degree of population subdivision on the basis of phylogenetic analyses alone. Much more information about population history can be gleaned from DNA sequences than by simply considering what is found on the most basal nodes of a gene tree. In this chapter, I describe the results of an entire class of population genetic analyses that can detect aspects of population history and gene flow that can augment phylogenetic reconstructions. Some of these analyses rely on the accurate inference of a phylogenetic tree, while others do not thus freeing them from tree-based assumptions about population history.

The entire chimpanzee HV-I sequence database (n=254) is used in this chapter to provide a detailed description of the geographical distribution of genetic diversity, and to directly address questions 2 and 3 outlined in Chapter 1. Specifically, these analyses are used to identify the most likely phylogeographic barrier(s) in Nigeria and Cameroon, and to examine how effectively and for how long phylogeographic barriers have influenced chimpanzee distribution. Whenever possible, measures of genetic diversity of HV-I sequences from chimpanzees in Nigeria and Cameroon are directly compared to the same, or similar, measures of genetic diversity calculated for chimpanzee populations from other areas. The first sections of this chapter provide a detailed description of gene flow between chimpanzee populations in Nigeria and Cameroon from a comparative perspective. The final sections of this chapter attempt to model aspects of the recent demographic histories of different regional chimpanzee populations, and provide rough estimates the relative age of each regional chimpanzee population.

Cladistic analysis of gene flow

Theory and methods

Slatkin and Maddison (1989, 1990) presented a method to detect restrictions to gene flow and population isolation. Using this model, it is possible to reconstruct the number of migration events needed to explain the spatial distribution of genetic diversity. However, this analysis is completely dependent upon inferring the correct phylogenetic tree. In this approach, location is coded as an unordered multistate character. The evolution of that character is traced along a tree reconstructed from genetic information. The minimum number of character state changes reconstructed along the tree represents the minimum number of inferred migration events, s. The value of s obtained from the reconstructed tree can be compared to s values derived from a large number of randomly generated trees of equal sample number to evaluate the statistical significance of the observed s value. If the observed number of migration events is lower than 95% of the randomly generated s values, the null hypothesis may be rejected and the conclusion drawn that gene flow is restricted in the sample (Slatkin and Maddison 1989; Hudson 1990; Slatkin and Maddison 1990; Hudson, Slatkin, and Maddison 1992; Slatkin 1994b).

Results and interpretation

The Slatkin and Maddison technique was applied to the trees in Figures 3.1-3.6, and to subsets of those trees. For the entire data set a single multistate, unordered character was assigned to each sequence denoting regional provenance. Regional levels were defined as (1) Upper Guinea (including samples from western Nigeria), (2) eastern Nigeria and western Cameroon north of the Sanaga River, (3) western equatorial Africa beginning in Cameroon south of the Sanaga River and (4) eastern Africa including all samples from chimpanzees east of the Ubangi River. Within the Nigeria-Cameroon HV-I sequence database, two multistate unordered characters were assigned, one character denoting whether the sample was collected from western Nigeria (Figure 2.2, map location 1), eastern Nigeria and western Cameroon north of the Sanaga River (Figure 2.2, map locations 3-12) or from Cameroon south and east of the Sanaga River (map locations 13 –16) and one character denoting the sampling location for each sequence. Null distributions of each character were traced along 1,000 random trees using the computer program MacClade (Maddison and Maddison 1996).

Figure 4.1 compares empirical *s* values for the entire data set from the parsimony, neighbor-joining and maximum likelihood trees depicted in Figures 3.1-3.6. For the regional provenance character, observed *s* values were lower than all values generated in the simulated null distribution. This observation implies that gene flow is significantly restricted at the regional level ($p < 0.05$).

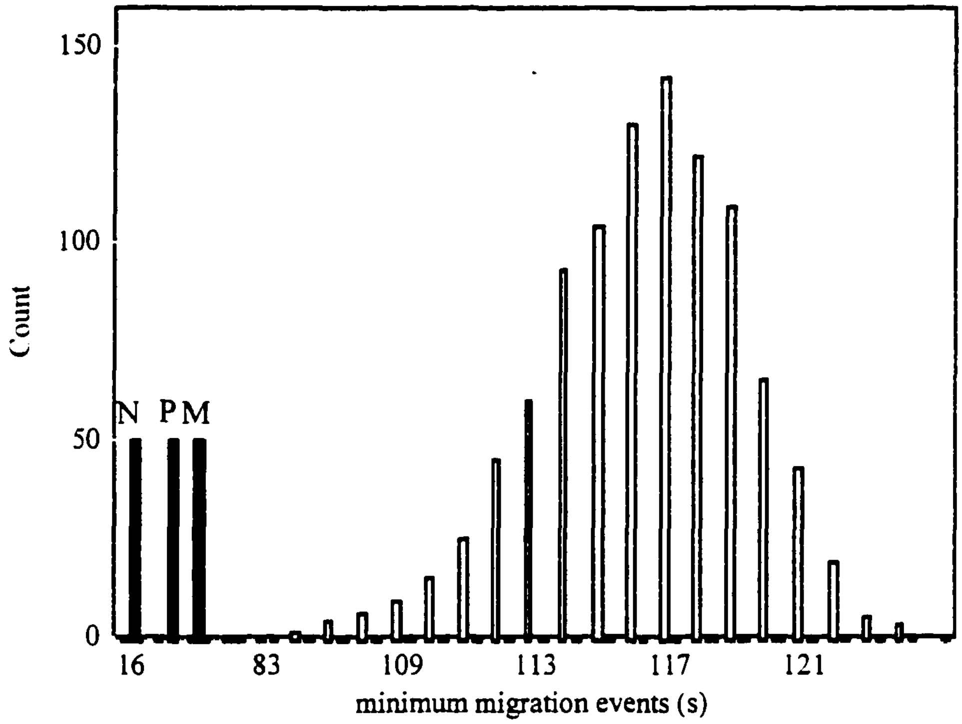


Figure 4.1. Minimum number of inferred migration events among chimpanzees across Africa. This histogram is based on phylogenetic reconstructions and 1,000 randomly-generated trees for 254 pan-African HV-I DNA sequences. Each haplotype was assigned a regional character denoting provenance. The evolution of this character was traced along the 1,000 random trees generated in MacClade (Maddison and Maddison 1996). Open bars represent the null distribution of the regional provenance character. The black bar labeled, N, represents the empirical value of the regional provenance character traced along the jackknife neighbor-joining tree depicted in Figure 3.5. The black bar labeled, P, represents the empirical value of the regional provenance character traced along the parsimony tree depicted in Figure 3.1. The black bar labeled, M, represents the empirical value of the regional provenance character traced along the maximum likelihood tree shown in Figure 3.3. See text for divisions of the regional provenance character.

Figure 4.2 depicts empirical *s* values and null distributions for sequences derived from samples collected within the study area. Samples were divided in two ways. The top panel of Figure 4.2 illustrates empirical *s* values for most likely regional division of the HV-I sequence database: Nigeria and western Cameroon north of the Sanaga River and Cameroon south of the Sanaga River. Samples from western Nigeria were excluded from this histogram. Empirical *s* values were lower than 95% of the null distributions for this character, implying significantly restricted gene flow across the Sanaga River ($p<0.05$). The bottom panel of Figure 4.2 depicts empirical *s* values the location character. At the population level, empirical *s* values were lower than for all *s* values obtained from the null distributions generated for that character. This observation implies that gene flow is significantly restricted at the regional and population levels among chimpanzees in Nigeria and Cameroon ($p<0.05$).

Gene flow and the distribution of shared haplotypes

Goldberg (1996) reported that when the coalescent model was applied to HV-I sequence phylogenetic reconstructions, gene flow was significantly restricted among chimpanzees in eastern Africa at the population and regional levels. Similarly, when the coalescent model was applied to sequences representing chimpanzees in Upper Guinea, gene flow was also restricted (Gagneux 1998). Given the quantitative similarities between cladistic analyses of gene flow, the qualitative differences in the extent of gene flow between regional chimpanzee populations are unclear. Qualitative differences in gene flow can be assessed between regional populations by examining the geographical distribution of shared sequences between populations.

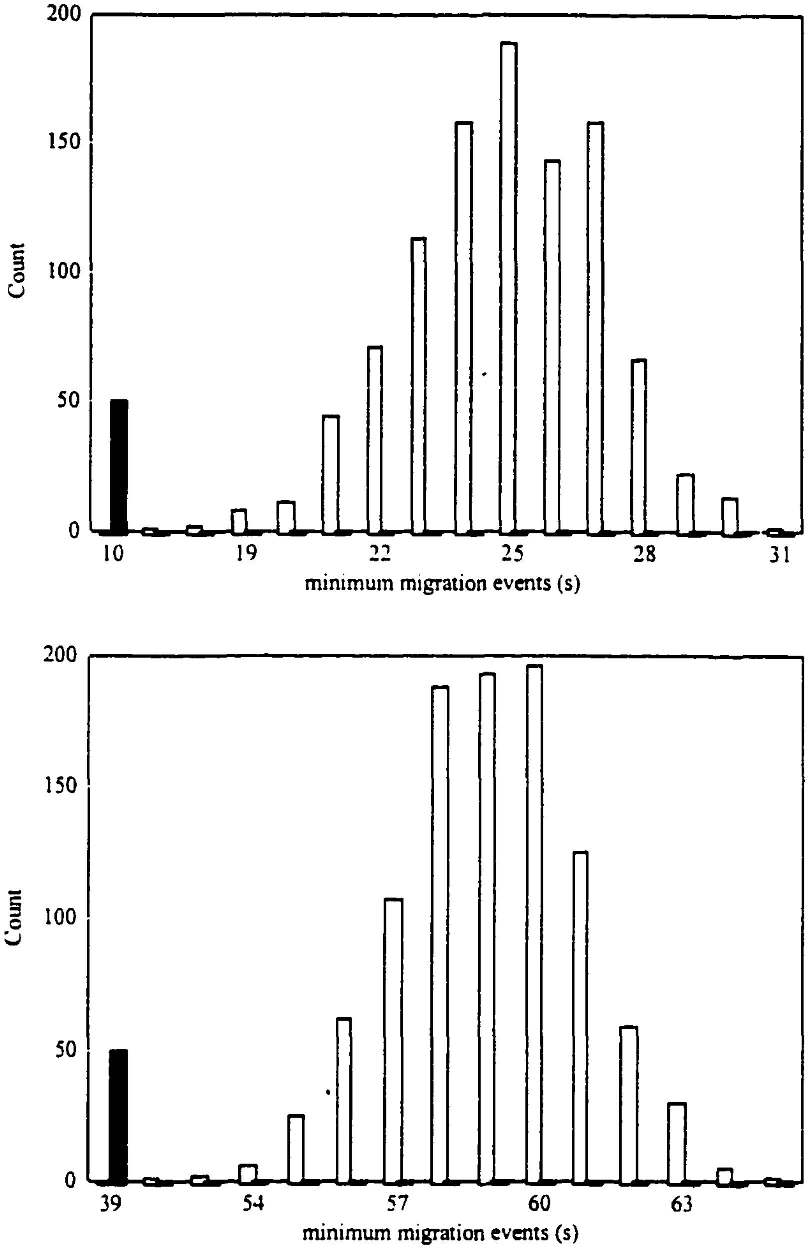


Figure 4.2. Minimum number of inferred migration events among chimpanzees in Nigeria and Cameroon. This histogram is based on phylogenetic reconstructions and 1,000 randomly-generated trees for 78 unique HV-I DNA sequences from samples gathered in Nigeria and Cameroon. The evolution of the regional and location characters were traced along the 1,000 random trees generated in MacClade (Maddison and Maddison 1996). Open bars represent the null distribution of the regional provenance character. The black bar represents the empirical values of *s* traced along the phylogentic reconstructions in Figures 3.1-3.6. The top panel shows empirical and null *s* values for a regional division of haplotypes divided along the Sanaga River in Cameroon. The bottom panel shows empirical and null *s* values for haplotypes divided by sampling location.

Endemic taxa exist exclusively within a restricted geographic range, while taxa with cosmopolitan distributions are eurytopic (Cox and Moore 1993). Goldberg (1996, 1998) applied these concepts to HV-I sequence data obtained from chimpanzees in eastern Africa. He considered a sequence endemic if it occurred only in one sampling location. Conversely, he considered sequences that appeared in more than one sampling location eurytopic. Comparisons between the eurytopic distributions of sequences between different regional chimpanzee populations reveal some interesting properties about the relative differences between the regional chimpanzee populations.

Three of the 79 unique HV-I sequences identified in the Nigeria-Cameroon database represent composite haplotypes shared by individuals within groups or between populations. Only one haplotype is eurytopic (GONCOMP01, n = 4). This haplotype was detected at Akoh Zanto, Ngel Nyaki Forest Reserve and Gashaka Gumti National Park (Figure 2.2, map locations 5 – 9). These sampling locations are situated in an outlying region of the Cameroon Highlands in Nigeria, and are separated by a maximum span of 200 km. The other two shared haplotypes occurred only within sampling locations in Cameroon, one at Campo Forest Reserve (GONCOMP02, n = 3, map location 15) and one at Dja Biosphere Reserve (GONCOMP03, n = 3, map location 16).

These observations contrast with similar studies. Goldberg (1996) found evidence of eurytopic haplotypes from sampling locations separated by 583 km in eastern Africa; while Gagneux (1998) identified eurytopic haplotypes between populations separated by over 1,000 km among chimpanzees in Upper Guinea. The differences in the distribution of eurytopic haplotypes between populations in these regions suggest that although gene flow may be restricted within each regional population, there are qualitative differences in the gene flow process, and probably, in the paleodemographic histories of chimpanzees in different regions.

F-Statistics and Analysis of Molecular Variation

Theory and methods

The Analysis of Molecular Variation, or AMOVA, is derived from Wright's (1951) F_{ST} model. Wright's model estimates F-statistics that measure the deviation of genotype frequencies in subdivided populations from a null model of panmixia. These statistics, or fixation indices, are usually divided into three measures of genetic diversity: within populations (F_{IT}), between populations within regions (F_{IS}) and between regions (F_{ST}). Positive fixation indices indicate departures from Hardy-Weinberg expectations. Fixation indices that approach zero indicate an absence of genetic structure. This model has been adapted to a wide range of genetic systems and has been refined to produce reliable results under different evolutionary assumptions, with finite sample sizes and with multiple loci (e.g., Cockerham 1969; Nei 1977; Weir and Cockerham 1984; Nei 1987; Lynch and Crease 1990; Slatkin 1994a; Goldstein *et al.* 1995; Slatkin 1995).

Following adaptations of Wright's F_{ST} model by Cockerham (1969), AMOVA also partitions total genetic variation into variation a) within populations, b) among populations within regions, and c) among regions. In the last partition, regions are exclusive groups of populations that are defined by *a priori* criteria, such as the regional divisions inferred by the phylogenetic reconstructions. Excoffier, Smouse and Quattro (1992) incorporated information about molecular distances among haplotypes into the AMOVA approach. Traditional F-statistics assume equidistance among alleles, basing the calculation of fixation indices on the absolute number of pairwise differences between sequences. In the AMOVA approach, "hidden" genetic variation may be revealed because a specific model of sequence evolution can be incorporated into the calculation of fixation indices. In the AMOVA model, "ϕ-statistics" are treated as molecular analogues of traditional fixation indices. In this approach ϕ_{ST}, ϕ_{SC} and ϕ_{CT} are used as molecular analogues of fixation indices within populations, among populations within regions and among regions, respectively. At each of these levels, ϕ-statistics are defined as:

4.1) $$\phi_{CT} = \frac{\sigma_a^2}{\sigma_T^2}$$

4.2) $$\phi_{SC} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2}$$

4.3) $$\phi_{ST} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_T^2}$$

where the total molecular variance (σ_T^2) is the sum of the variances due to differences among haplotypes within populations (σ_c^2), variance due to differences between haplotypes in different populations within regions (σ_b^2), and variance due to differences among populations (σ_a^2) (Excoffier, Smouse, and Quattro 1992; Schneider *et al.* 1997).

Results and interpretation

The AMOVA approach was applied to samples from Nigeria and Cameroon and to samples across Africa using the computer program Arlequin (Schneider *et al.* 1997). AMOVA's were generated for several different divisions of the data. The iteration with lowest variance and highest ϕ_{CT} value was the major population division at the Sanaga River in Cameroon that was inferred from the gene trees depicted in Figures 3.1 to 3.6. Samples from Nigeria and Cameroon were divided into two groups: eastern Nigeria and western Cameroon north of the Sanaga River or Cameroon south of the Sanaga River. These HV-I sequences were further subdivided into populations by sampling location, assuming that a different population is represented by each sampling location. Across Africa, chimpanzee sequences were divided into regions 1-4 described on p. 113. Sample sizes of sequences from chimpanzees in western Nigeria were too small to permit statistical testing, and were excluded from all AMOVA analyses.

AMOVA results for the Nigeria-Cameroon sequences are presented in Table 4.1. Two matrices of inter-allelic distance were analyzed using the computer program Arlequin (Schneider *et al.* 1997). The first matrix describes Euclidean distances among haplotypes, inferred from the number of nucleotide differences between pairs of sequences with regard to the nature of those differences. This phenetic matrix is called the "D1", or haplotypic, matrix. The second matrix corresponds to Excoffier, Quattro and Smouse's (1992) "D2" matrix in which all haplotypes differ by an equal genetic distance, regardless of the varying amounts of nucleotide sequence difference. The resulting "multiallelic" ϕ-statistics reduce to traditional F-statistics. Tables 4.2 and 4.3 have been included for comparison. Table 4.2 lists AMOVA values for eastern African

Table 4.1. Hierarchical analysis of molecular variance (AMOVA) for chimpanzee HV-I DNA sequences from Nigeria and Cameroon

Populations regionally divided by the Sanaga River*

Variance Component		D1, Haplotypic Variance†				D2, Multiallelic Variance‡			
		Observed Partition				Observed Partition			
		Variance	% Total	P^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions	σ^2_a	5.732	33.10	<0.001	$\phi_{CT} = 0.331$	5.048	31.65	<0.001	$\phi_{CT} = 0.317$
Among populations/regions	σ^2_b	2.694	15.33	<0.001	$\phi_{SC} = 0.229$	2.411	15.12	<0.001	$\phi_{SC} = 0.221$
Within populations	σ^2_c	8.930	51.57	<0.011	$\phi_{ST} = 0.484$	8.489	53.23	<0.011	$\phi_{ST} = 0.468$

*Samples were regionally divided by location into eastern Nigeria plus western Cameroon north of the Sanaga river (Figure 2.2, map locations 3-12) and Cameroon south of the Sanaga River (map locations, 13-16). Samples from western Nigeria were excluded from this analysis because they may be part of an Upper Guinea chimpanzee lineage.

†A Tamura and Nei (1993) distance matrix among haplotypes was used to calculate haplotypic variances

‡A pairwise distance matrix assuming all haplotypes were equidistance was used to calculate multiallelic variances

Table 4.2 Hierarchical analysis of molecular variance (AMOVA) for chimpanzee HV-I DNA sequences from eastern Africa

Populations regionally defined as forests*

Variance Component		D1, Haplotypic Variance†				D2, Multiallelic Variance‡			
		Observed Partition				Observed Partition			
		Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions	σ^2_a	0.076	1.88	0.131	$\phi_{CT} = 0.019$	0.000	0.00	0.879	$\phi_{CT} = 0.000$
Among populations/regions	σ^2_b	0.561	13.87	<0.001	$\phi_{SC} = 0.141$	0.051	10.23	<0.001	$\phi_{SC} = 0.102$
Within populations	σ^2_c	3.409	84.25	<0.001	$\phi_{ST} = 0.158$	0.447	89.77	<0.001	$\phi_{ST} = 0.102$

*Data from Goldberg (1996). The data were divided into eastern forests (Uganda, Rwanda and Tanzania) vs. Democratic Republic of Congo forests (n = 255)

†A distance matrix among haplotypes uncorrected for multiple substitutions was used to calculate haplotypic variances

‡A pairwise distance matrix assuming all haplotypes were equidistant was used to calculate multiallelic variances

Table 4.3. Pairwise ϕ_{ST} values of HV-I sequences in Upper Guinea

Populations defined as sampling locations*

	Tai	A	B	Nimba	Comoe	Bossemattie	Dagbego	Tenkere	Solo
Tai		(5)	(5)	(220)	(460)	(400)	(190)	(650)	(830)
A	0.04993		(5)	(225)	(460)	(400)	(185)	(650)	(830)
B	0.03856	0.00118		(200)	(455)	(395)	(185)	(655)	(830)
Nimba	0.05557	**0.1744**	0.20512		(500)	(540)	(400)	(460)	(630)
Comoe	**0.18033**	**0.4722**	**0.47484**	**0.22902**		(260)	(460)	(860)	(820)
Bossemattie	-0.0088	0.1047	0.06674	0.02061	**0.25302**		(290)	(980)	(1020)
Dagbego	-0.00169	0.1723	**0.15801**	0.09326	**0.25431**	-0.00539		(840)	(955)
Tenkere	0.01952	0.00577	0.0271	0.09862	**0.31593**	0.07364	**0.1233**		(380)
Solo	0.01415	0.1535	0.08027	0.06234	**0.25293**	-0.03389	-0.01534	**0.11365**	

*Data taken from Gagneux (1998); pairwise ϕ_{ST} shown in lower diagonal; geographic distance between populations given in km shown in upper diagonal; bold values are significant at $p < 0.05$ as tested in 1000 random permutations chimpanzees taken from Goldberg (1996). Table 4.3 lists ϕ_{ST} values for chimpanzees in Upper Guinea taken from Gagneux (1998).

Several different regional divisions of chimpanzees within the study area were computed in the AMOVA framework. However, when samples were divided along the Sanaga River variances were the lowest. Consequently, this subdivision is the most likely regional division of chimpanzees within Nigeria and Cameroon. The distribution of genetic variation among chimpanzees throughout Nigeria and Cameroon region is greatly different from chimpanzees elsewhere. Regional variation among the Nigeria-Cameroon sample accounts for 33.10% (D_1 matrix) and 31.65% (D_2 matrix) of the total variation. Among eastern African chimpanzees, divided regionally between Democratic Republic of Congo forests and eastern forests (Uganda, Rwanda and Tanzania), regional variation accounts for 1.88% (D_1 matrix) and 0% (D_2 matrix) of the genetic diversity detected in the sample. Genetic diversity at the population level was different as well, accounting for 15% of the variation detected among samples from Nigeria and Cameroon and 10-14% of the variation detected among chimpanzees in eastern African forests. Only half of the genetic variation detected in Nigeria and Cameroon is attributable to variation within populations. Conversely, 85-90% of the genetic variation is attributable to variation within populations among eastern African chimpanzees. Possible regional divisions among chimpanzees in Upper Guinea were unavailable. However, ϕ_{ST} values for chimpanzee populations in that region range from –0.00169 to 0.4784, with a mean haplotypic ϕ_{ST} value of 0.119. Among chimpanzees in eastern Africa, the haplotypic ϕ_{ST} value is 0.131. In contrast, samples from Nigeria and Cameroon reveal a much higher haplotypic ϕ_{ST} value of 0.484.

These findings suggest that significant genetic subdivision exists on either side of the Sanaga River between populations in eastern Nigeria and western Cameroon north of the Sanaga River and those in Cameroon south of the Sanaga River. Moreover, the genetic differences between samples on either side of the Sanaga River account for more of the genetic diversity among chimpanzees than does any other regional division within or outside the study area. These findings are markedly different for other highly-vagile species which generally have very low fixation indices; and is more similar to fixation indices of smaller animals that tend to disperse over small distances (Allendorf 1983; Nei 1987; Avise 1994).

Null distributions of AMOVA variance components were determined by 16,000 random permutations of the data to test the hypothesis that the observed variance components resulted from sampling error, rather than from real population substructure (Excoffier, Quattro, and Smouse 1992). This approach is nonparametric, making few *a priori* distributional assumptions about the data. Observed variances were compared to the simulated null distributions to assess the probability of obtaining a variance as extreme as the empirical value by chance alone. If the observed variance components are higher or lower than 95% of the simulated null distributions, then the null hypothesis may be rejected that the detected population substructure was the result of sampling error.

Results of these distributions are shown in Figure 4.3. Open bars represent the 16,000 simulated null distributions. Closed bars represent the empirical values of the variance components. Information about molecular distances among haplotypes did not change the results found in the standard multiallelic approach. Chimpanzees in Nigeria and Cameroon show a significant excess of genetic variation both at the regional

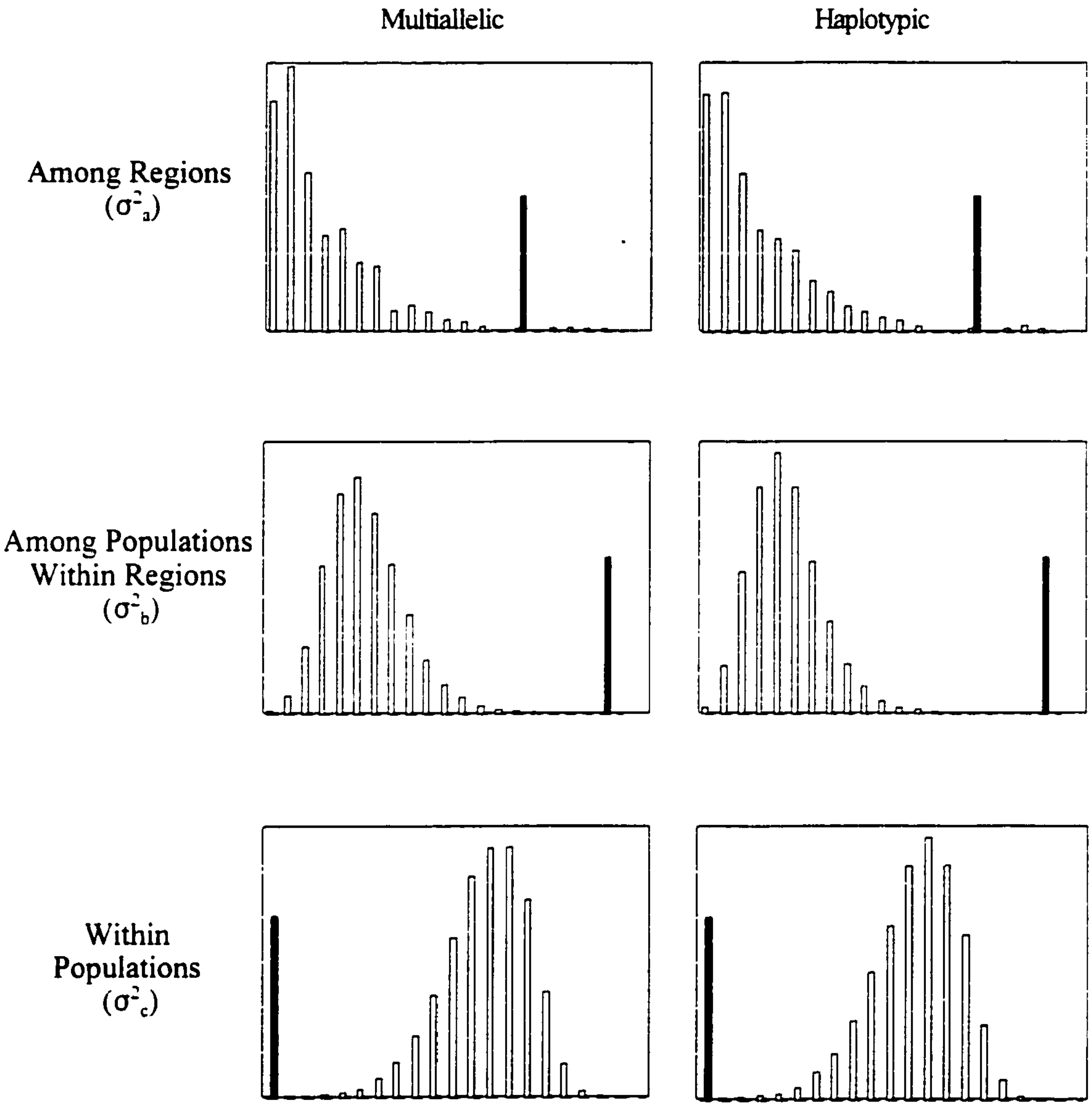


Figure 4.3. Null distributions of multiallelic and haplotypic variance components at three hierarchical levels. Open bars show the distribution of variances generated by 16,000 random permutations of a) populations across regions, b) populations within regions, and c) individuals within populations. 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: populations in eastern Nigeria and western Cameroon north of the Sanaga River and populations in Cameroon south and east of the Sanaga River.

($p < 0.001$) and population levels ($p < 0.001$). Within populations, however, chimpanzees in Nigeria and Cameroon show a statistically significant lack of within-population genetic diversity ($p < 0.01$) compared with what the null model predicts. This lack of intra-populational divergence suggests that individual chimpanzees carry most of the genetic diversity that exists in the species. Most of the genetic differentiation at this locus, however, is attributable to differences detected between individuals and populations divided by the Sanaga River.

Migration rates

N*m* estimates the absolute numbers of migrants per generation between populations, and is the product of effective population size and the absolute number of migrants per generation. N and *m* are difficult to estimate separately. Multiple genetic systems must be analyzed to obtain accurate migration estimates (Slatkin 1987; Avise 1994; Slatkin 1994a). However, Takahata and Palumbi (1985) demonstrated that the product of N and *m* could be obtained from F_{ST} estimates of mitochondrial loci by:

4.4) $$Nm = \frac{1 - F_{ST}}{2F_{ST}}$$

Table 4.4 lists estimates of the product of N and *m* for chimpanzees in Nigeria and Cameroon compared to chimpanzees in eastern Africa and those in Upper Guinea. Although the values listed in Table 4.4 are very rough estimates, it is clear that fewer migrants (1.1) per generation have been exchanged between chimpanzee populations on

Table 4.4. Comparative migration rates (N*m*) between regional chimpanzee populations

	Estimated Migration Rate (N*m*)	
Region	Between Regions	Between Populations within Regions
Eastern Nigeria and Cameroon	1.1 (across the Sanaga River)	1.8
Eastern Africa†	25.8 (between regional forests)	3.0
Upper Guinea‡	n/a	Mean = 6.02*

†Data taken from Goldberg (1996).
‡Data taken from Gagneux (1998).
*Regional divisions were not given in Gagneux (1998). This value represents mean migration rates between pairs of populations in Upper Guinea.

either side of the Sanaga River. Conversely, 25 times that number of migrants have been exchanged between chimpanzee populations inhabiting forests in eastern Africa; while 6 times as many migrants have been exchanged between chimpanzee populations in Upper Guinea. Theoretically, only one migrant per generation is sufficient to prevent population subdivision by genetic drift alone (Nei 1987; Avise 1994; Weir 1996). This study suggests, however, that a significant genetic discontinuity exists in the vicinity of the Sanaga River in central Cameroon, perhaps mediated by genetic drift or by other less obvious evolutionary constraints. Clearly, chimpanzee migration across the Sanaga River has been much less extensive than between populations and communities in other regions of Africa.

Regional comparisons of genetic diversity

Slatkin's linearized F_{ST}

The previous analyses suggest that there are varying degrees of genetic subdivision within regional chimpanzee populations. However, the degree of division between regional chimpanzee populations remains unclear. Slatkin (1991, 1993, 1994a) showed that F_{ST} estimates can provide more detailed information about population divergence and isolation by distance. This approach is formally referred to as the linearized F_{ST} model. In this model, Slatkin evaluates a series of haploid populations of size N that diverged τ generations ago from a single ancestral population. Under these conditions, Slatkin (1991, 1994a) proposed that F_{ST} can be expressed in terms of coalescent times t_1, which is the mean coalescent time of two genes drawn from two different "sibling" populations. t_0 is the mean coalescent time of two genes drawn from the same population. Slatkin (1991, 1994a) expressed population pairwise F_{ST}'s as:

4.5) $$F_{ST} = \frac{t_1 - t_0}{t_1}$$

Because t_0 is equal to N generations (Hudson, 1990), and t_1 is equal to $\tau + N$ generations, the above expression reduces to $F_{ST} = \tau/\tau+N$. Therefore, the ratio:

4.6) $$D = \frac{F_{ST}}{1 - F_{ST}}$$

is equal to τ/N, and is proportional to the divergence time between populations. Small values of D reflect more recent associations between pairs of populations. Larger values of D indicate more increased genetic differentiation between pairs of populations and more ancient relative coalescent events.

Results and interpretation

The linearized F_{ST} approach was applied to the full data set, divided into major regional populations. Chimpanzee populations were divided into the four divisions listed on P. 113. All calculations were performed using the computer program Areliquin (Schneider *et al.* 1997).

Results are presented in Table 4.5. Values in the upper diagonal represent Slatkin's (1991) linearized F_{ST} values, D. Populations on the extreme western and

Table 4.5. Pairwise and linearized F_{ST} estimates* between regional chimpanzee populations

Region	Upper Guinea	Nigeria and Cameroon north of Sanaga R.	Western Equatorial Africa	Eastern Africa
Upper Guinea (including western Nigeria)†		0.908 (0.792)	1.139 (1.058)	3.146 (2.858)
E. Nigeria and W. Cameroon north of Sanaga R.	0.475 (0.442)		0.680	1.876
Western Equatorial Africa (includes Cameroon south of Sanaga R.)	0.533 (0.514)	0.405		0.668
Eastern Africa	0.759 (0.741)	0.653	0.401	

*Pairwise F_{ST} estimates are shown in the lower diagonal. Linearized F_{ST} estimates (Slatkin, 1991) are shown in the upper diagonal. †Numbers to the left of the parentheses were calculated excluding samples from western Nigeria. Numbers in parentheses were calculated including samples from western Nigeria.

eastern edges of chimpanzee distribution are the most different, with varying degrees of increasing similarity moving toward the interior portions of Africa. The results are more interesting when *D* values are compared between regional western, western equatorial and eastern African populations. The pairwise *D* value between the regional chimpanzee populations in western Africa is 0.908 excluding samples from western Nigeria, and 0.792 including them. Given the ambiguous position of the western Nigerian samples in the phylogenetic reconstructions presented in Chapter 3, the differences between these *D* values may reflect the unique sequence composition of the samples from Ise Forest Reserve. The pairwise *D* value between the western equatorial and eastern African regional chimpanzee populations (0.668) is a lower than the *D* values calculated for chimpanzees in western Africa. These *D* values suggest that the regional chimpanzee populations in western Africa are more genetically distinct from each other than the western equatorial and eastern African populations are from one another. Moreover, because *D* is proportional to a population's coalescent time, the regional chimpanzee populations in western Africa probably are much older than are those in eastern Africa. The pairwise *D* value for the eastern Nigeria and western Cameroon regional population and the western equatorial population is low ($D = 0.680$). This value may be lower than expected due either to recent but infrequent migrations across the Sanaga River or may reflect the mix of "western haplotypes" and " central haplotypes" found at Manb'ra (Figure 2.2, map location, 13), which is east of Mbaum River in Cameroon.

Spatial Autocorrelation

Theory and methods

AMOVA, F_{ST} analyses and migration estimates are useful summaries of genetic diversity and population substructure, but generally make several *a priori* assumptions about the demographic structure of the population or species of interest. Spatial autocorrelation is defined as the association of the values of one variable with the values of the same variable at all other localities (Sokal and Oden 1978a; Sokal and Oden 1978b). This approach has been used widely in ecological studies (reviewed in: Koenig 1999), and is particularly useful in analyzing spatial aspects of genetic diversity (Sokal and Oden 1978a, 1978b; Oden 1984; Barbujani 1987; Bertorelle and Barbujani 1995). Unlike other single-point estimators of genetic diversity, spatial autocorrelation gives a detailed description of gene frequency variation through space, and is independent of *a priori* assumptions about the underlying population structure (Barbujani 1987). Ideally, an autocorrelation index for DNA analysis should measure whether, and to what extent, individual DNA sequences resemble the sequences sampled at different locations.

Bertoelle and Barbujani (1995) proposed two autocorrelation indices, or AIDAs: a product-moment coefficient, analogous to Moran's *I*, and a distance-like coefficient, analogous to Geary's *C* (Sokal and Oden 1978a). Only the derivative of Moran's *I*, *II*, is presented in this chapter. This AIDA is given by:

$$II = \frac{\sum_{i=1}^{n-1} \sum_{j \geq i}^{n} (p_{ik} - \overline{p}_k)(p_{jk} - \overline{p}_k)}{W \sum_{i=1}^{n} \sum_{k=1}^{S} (p_{ik} - \overline{p}_k)^2} \tag{4.7)}$$

where W is the number of pairwise comparisons in a given distance class. The binary variable, p, represents the identity (1) or nonidentity (0) of a nucleotide along a vector of S sites representing one of n individual sequences in a population sample. Using this equation, each DNA sequence is treated as an independent observation, as opposed to each population as independent observation. Moreover, it is easy to assign different weights to different sequences or regions within sequences or to incorporate a TS:TV ratio (Bertorelle and Barbujani 1995).

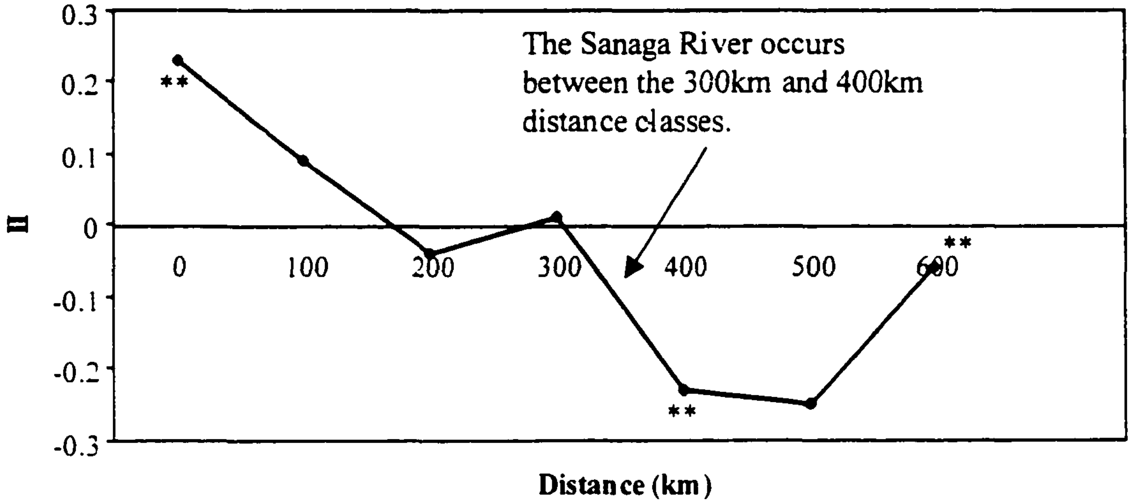
A spatial autocorrelagram plots the autocorrelational strength of the sequences as a function of geographic distance (Smouse and Peakall 1999). A positive autocorrelation denotes similarity between sequences and is shown by positive II values. Conversely, a negative autocorrelation denotes dissimilarity between sequences and results in II values at the other extreme. Clinal trends are indicated by autocorrelagrams that decrease monotonically from positive to negative with distance. Autocorrelations yielding more complex patterns may indicate varying kinds and degrees of genetic variation across space. The statistical significance of an autocorrelagram is inferred by comparison of the observed autocorrelation index to a null distribution generated by a large number of random permutations of the data (Bertorelle and Barbujani 1995).

Results and interpretation

In order to perform the AIDA test, the study area in Nigeria and Cameroon was divided into six distance classes, using a greatest circle criterion defined by Bertoelle and Barbujani's (1995) AIDA program. The center of the study area was determined by drawing the smallest possible square that included all sampling locations on a map of the study area. The center point of the square was defined as the center of the study area in Nigeria and Cameroon. Distance classes were determined by drawing six concentric circles of increasing diameter in equal increments inside the square. All sampling locations included within the same circle were counted as part of the same distance class. Using this method, the center of the study area was defined as roughly 50 km southwest of Akoh Zanto (map location, 5). Analyses were performed on six distance classes of even width (100 km) for the entire study area. Statistical significance of the data was assessed using 1000 random permutations of the data.

Figure 4.4 illustrates an AIDA for the entire study area in Nigeria and Cameroon. At the zero distance class HV-I sequences are significantly similar. Between the 300 and 400 km distance class, HV-I sequences are significantly dissimilar. The Sanaga River occurs between the 300 and 400km distance classes; and these results probably reflect the genetic discontinuity present along the banks of the river. At the 500 to 600km distance classes, a significant increase in *II* values was detected. This increase is probably attributable the fact that only samples from Dja Biosphere Reserve (map location, 16) were included in this distance class.

Figure 4.5 depicts AIDA's for samples collected Nigeria and western Cameroon north of the Sanaga River (top panel) and Cameroon south and east of the Sanaga River



The Sanaga River occurs between the 300km and 400km distance classes.

Distance Class	Sample Location*
0-100	5
100-200	3, 4, 6, 10
200-300	7, 11, 12
300-400	8, 9, 13, 14
400-500	1, 15
500-600	16

*Sampling locations shown in Figure 2.2

Figure 4.4. Spatial autocorrelagram based on DNA sequence data from Nigeria and Cameroon. Asterisks indicate significant AIDA's between different distance classes. Significance was determined by 1,000 random permutations of the data. The data were analyzed using the computer program AIDA (Bertorelle and Barbujani 1994).

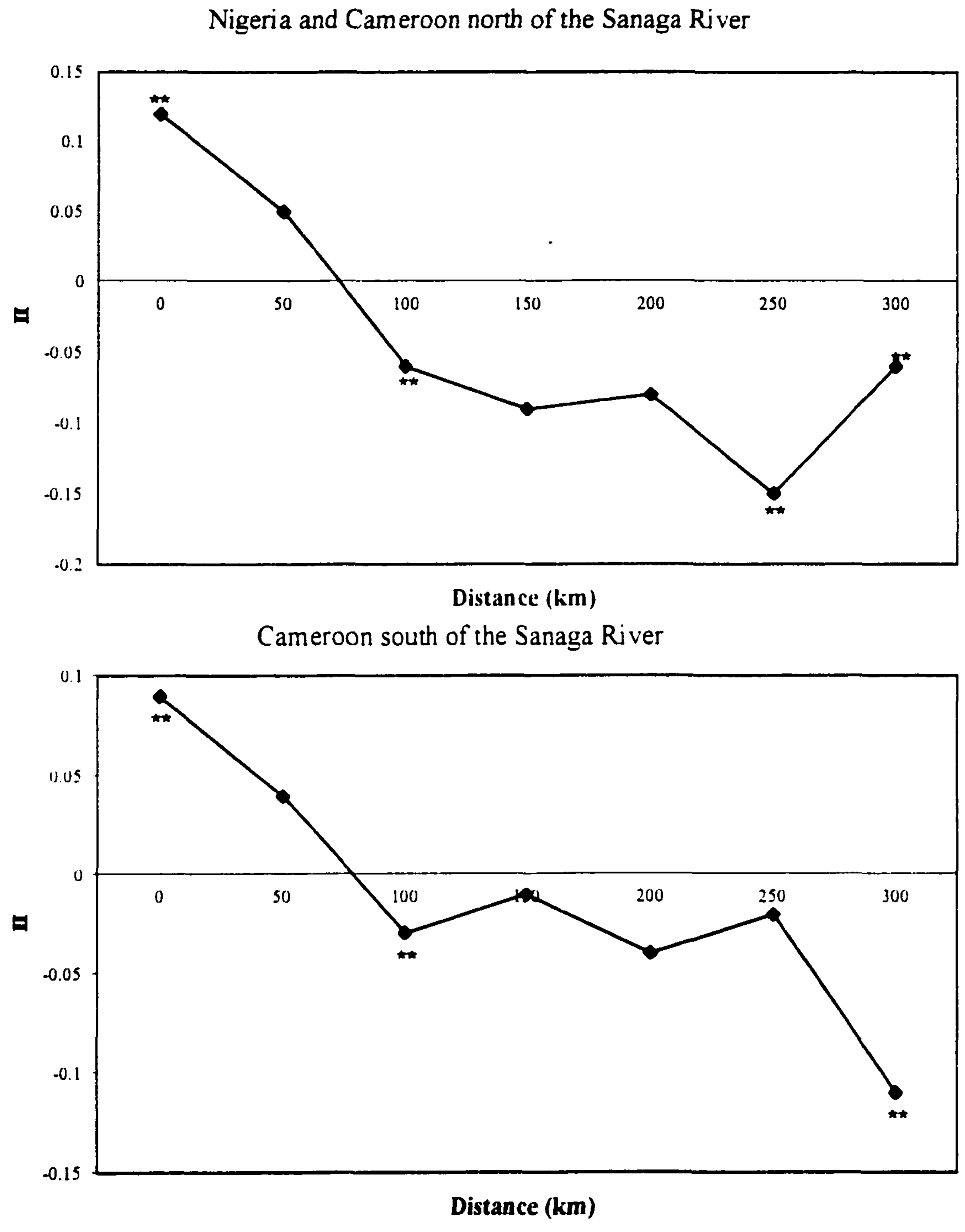


Figure 4.5. Spatial autocorrelagrams based on HV-I DNA sequence data divided across the Sanaga River. The top panel shows the autocorrelagram for samples collected in Nigeria and Cameroon north of the Sanaga River. The bottom panel shows the autocorrelagram for sequences from samples collected in Cameroon south of the Sanaga River.

(bottom panel). The overall pattern for the divided data set is somewhat similar to the observed pattern for the full data set. Both AIDA's show a significant positive spatial autocorrelation at the zero distance class, and generally have negative and decreasing autocorrelation values at larger distance classes. However, the top panel of Figure 4.5 shows a slight but significant decrease in genetic similarity between the 200 and 250 km distance classes for locations in Nigeria and Cameroon north of the Sanaga River. This decrease in *II* values could be attributable to two haplotypes from samples collected at the Cross River National Park in southeastern Nigeria. In the phylogenetic reconstructions for this locus, two of the Cross River National Park samples (CRNP6 and CRNP12) consistently clustered within the central African clade.

The bottom panel of Figure 4.5 shows a significant decrease in sequence similarity between the 250 and 300 km distance classes. Manb'ra was included within this distance class. Sequences from Manb'ra are evenly divided into the western African and central African clades in the phylogenetic reconstructions presented previously. It seems likely that the decrease in sequence similarity at this distance class may reflect the presence of those "western haplotypes" in a predominantly "central haplotype" region.

Characterization of population expansion

The preceding analyses illustrate that extensive population subdivisions exist at the HV-I locus between and within regional chimpanzee populations, but provide little insight as to how these differences may have developed. It is not clear how or when population expansions occurred across Africa or in Nigeria and Cameroon. The analyses presented in this section attempt to model how and when chimpanzee population expansions may have occurred throughout Africa.

Mismatch and intermatch distributions

Theory and methods

Rogers and Harpending (1992) presented a method for inferring characteristics of recent demographic history by analyzing properties of mismatch distributions. Mismatch distributions have been used extensively to model aspects of human population expansion. Although these analyses primarily have been used to model human population expansion based on mtDNA control region sequence data (Rogers and Harpending 1992; Harpending *et al.* 1993; Harpending 1994; Sherry *et al.* 1994; Bertorelle and Slatkin 1995; Rogers 1995; Rogers and Jorde 1995; Schneider and Excoffier 1999), the analysis of mismatch distributions has been adapted to other genetic systems as well (Shriver *et al.* 1997; Jorde, Bamshad and Rogers 1998).

Mismatch distributions are histograms of modal pairwise genetic differences. Because mismatch distributions rely on modal pairwise genetic differences, they provide a more recent time window into evolutionary events than do tree-based analyses or summaries of genetic diversity that rely on maximal pairwise genetic differences. Moreover, mismatch distributions do not rely upon the accurate reconstruction of a phylogenetic tree. This property makes the analysis of mismatch distributions a good complement to tree-based analyses because the timeframe in which phylogenetically-informative mutations may arise is quite narrow among closely-related lineages (Rogers and Jorde 1995).

Figure 4.6 illustrates the relationship between mismatch distributions, population history and the coalescent events marking the divergence of the most basal clades on a gene tree (Rogers and Jorde 1995). In a mismatch distribution, the shape and position of the histogram "wave" represents population characteristics following the coalescence of the most divergent nodes on a phylogenetic tree (Rogers 1995). The crest of the wave occurs just prior to the population expansion event. The position and shape of the wave in the histogram is proportional to the age and demographic stability of the population since expansion.

Three parameters, derived from the theoretical mismatch distribution shown in Figure 4.6, define and delimit the wave-like form of the histogram:

4.8) $\theta_O = 2N_o u$

4.9) $\theta_1 = 2N_1 u$

4.10) $\tau = 2ut$, where $u = m_T\mu$

where N_O is the population size of breeding females preceding population expansion, N_1 is the post-expansion population size, t is the elapsed time since population expansion, u is the probability that a mutation strikes a nucleotide in the sequence each generation and m_T is the length of sequence under study (Rogers and Harpending 1992). The values of these parameters define the shape of the mismatch distribution. τ defines the location of the crest of the wave, and is directly proportional to the time at which the population underwent expansion.

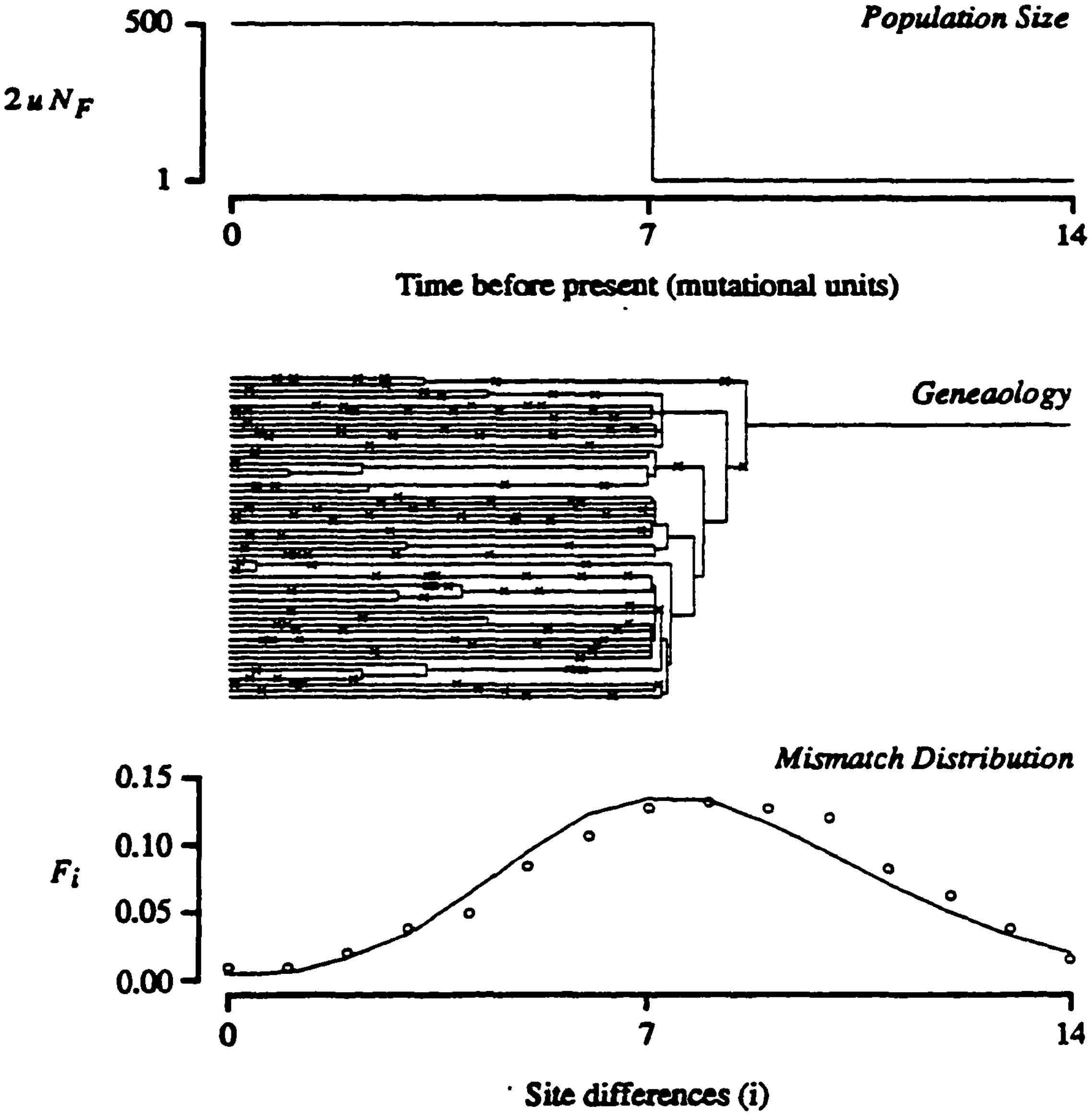


Figure 4.6. Relationship between population history, phylogeny and the mismatch distribution (redrawn from Rogers and Jorde 1995). The top panel shows a sudden expansion seven mutational units before the present. The middle panel shows the corresponding phylogeny, with crosses representing mutations. The lower panel shows the mismatch distribution for this theoretical data set. Mismatch distributions are histograms of modal pairwise site differences between sequences. Open circles represent empirical values; the solid line represents the null mismatch distribution for this simulated population.

Figure 4.7 illustrates the characteristic shapes of a theoretical continuous mismatch distribution. τ values define the location of the wave. Lower τ values indicate a population that has experienced a recent expansion. As τ increases, the wave moves tothe right indicating a more ancient population expansion (Rogers and Harpending 1992; Rogers and Jorde 1995). Harpending *et al* (1993) reported that mismatch distributions from expanding populations tend to be smoother than distributions generated from equilibrium populations; while demographically stable populations show a rough distribution.

There are several statistical models to evaluate the deviation of mismatch distributions from the continuous null model. Two were used during to analyze mismatch distributions at the HV-I locus. In order to discriminate between more recent expanding populations and more ancient demographically stable populations, Harpending (1994) proposed using the *ad hoc* statistic, *r* or Harpending's raggedness index, which is calculated as the sum of squared differences between sequential ordinal values in a mismatch distribution. Harpending's raggedness index (*r*) is calculated by:

4.11)

$$r = \sum_{i=1}^{d+1} (x_i - w_{i-1})^2$$

where *d* is the maximum number of observed differences between haplotypes, and the *x*'s are the observed relative frequencies of the mismatch classes (Rogers 1995). Low *r* values indicate a smooth unimodal distribution and a population history characterized by recent expansion; higher values indicate a rough multimodal distribution and a population

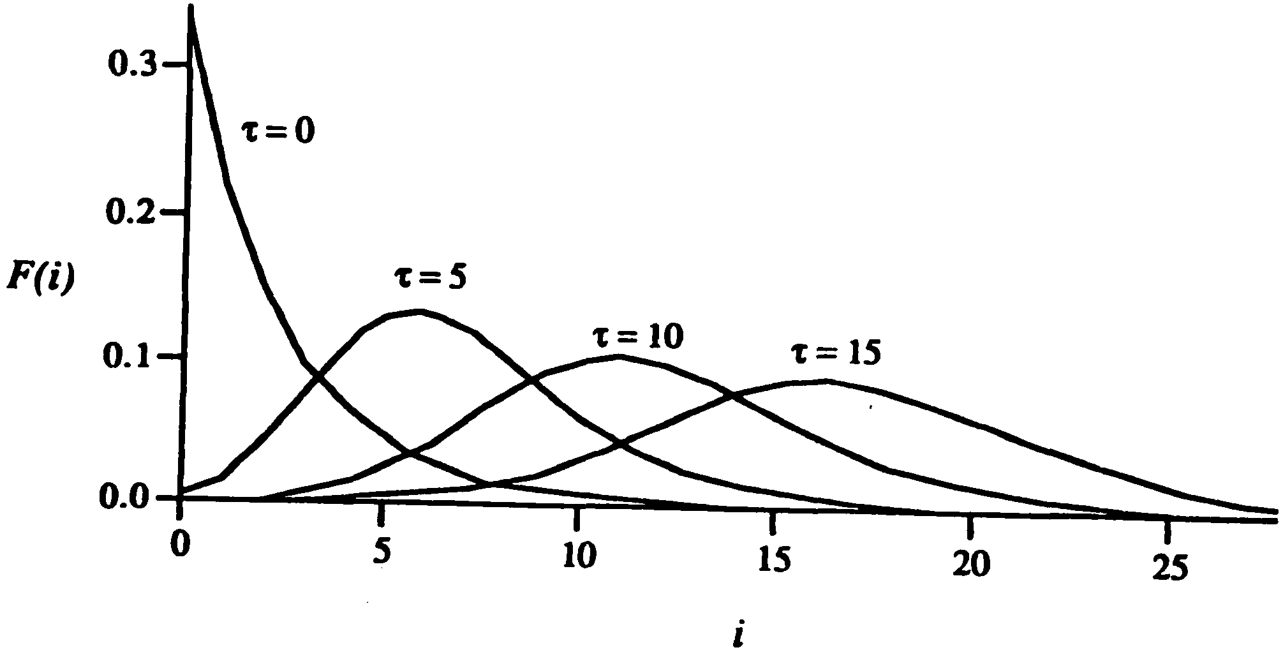


Figure 4.7. Relationship between the shape of the mismatch distribution and the parameter τ (redrawn from Rogers and Jorde 1995). The location of the crest of the wave is defined by τ, and is proportional to the time since population expansion. As the time since population expansion increases, the crest of the wave moves to the right, and τ values increase. At τ = 0, the simulated population is at complete demographic equilibrium. The absolute number of pairwise differences is represented by, *i*.

history characterized by demographic stability. Mismatch simulation models have shown that populations that have experienced rapid population growth show small *r* values of 0.012, while simulated distributions from equilibrium populations show *r* values of 0.26 (Harpending *et al.* 1993; Harpending 1994). The *r* statistic is robust to population subdivisioning, differing levels of migration between divided subpopulations, low-resolution data, small sample size and multiple substitutions (Rogers 1995; Rogers and Jorde 1995; Rogers *et al.* 1996; Schneider and Excoffier 1999). However, the shape of the histogram alone does not indicate whether the population significantly deviates from the null continuous mismatch model. Schneider *et al.* (1997) proposed a simple chi-square goodness of fit test to judge whether the observed mismatch distribution fits with the predicted population expansion scenario.

The modality of mismatch distributions characterizes aspects of population history. Marjoram and Donnelly (1994) showed that when two or more ancient lineages are present in a sample of DNA sequences, a bimodal distribution would be present in the mismatch histogram. These bimodal distributions are sometimes called intermatch distributions, and are a special case of the mismatch distribution (Harpending *et al.* 1993). Although the analytical methods for constructing the distributions are identical, the interpretations of them are different (Rogers and Jorde 1995). Bimodality may arise from two causes. First, the leading wave may reflect an expansion that occurred before the separation of the ancestral population. When a population expands and later splits into several subdivisions, the intermatch wave produced by the earlier expansion will continue to dominate the mismatch distribution within each subdivision. Secondly, the result is different when there is no wave in the original distribution. Then, the separation of a population produces a wave in the intermatch distribution that does not appear in the mismatch distributions. The characteristics of intermatch distributions have been used to construct various scenarios of human population expansion (Harpending *et al.* 1993; Marjoram and Donnelly 1994; Rogers and Jorde 1995).

Results and interpretation

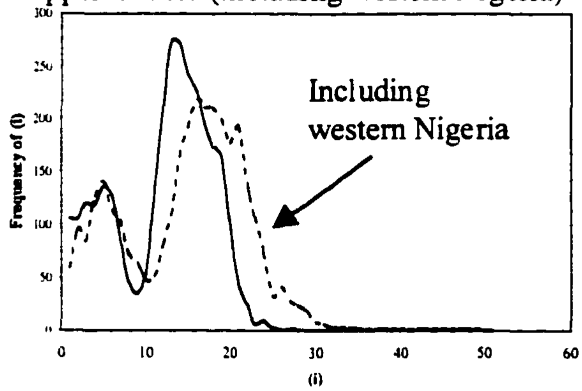
Mismatch distributions were calculated for different regional chimpanzee populations using the computer program Arlequin (Schneider *et al.* 1997). Figure 4.8 depicts the resulting histograms and associated statistics for samples collected during this study from Nigeria and Cameroon, from Upper Guinea (Gagneux 1998) and from eastern Africa (Goldberg 1996). The χ^2 tests for all mismatches were all highly significant; indicating that each regional population is not at demographic equilibrium.

The bottom panel of Figure 4.8 depicts the mismatch distribution for the comparable region of the HV-I locus in humans (data taken from Vigilant *et al.* 1991 and Goldberg 1996). The human data were included to highlight the similarities and the differences between different regional chimpanzee populations. Human mismatch distributions are often proposed as evidence of a recent human population expansion (Rogers and Harpending 1992; Harpending *et al.* 1993; Harpending 1994; Sherry *et al.* 1994; Bertorelle and Slatkin 1995; Rogers 1995; Rogers and Jorde 1995). Rogers and Jorde (1995) suggested that the mismatch distribution for the human HV-I locus corresponds to a human population expansion between 33 and 150kya.

The third panel of Figure 4.8 illustrates the mismatch distribution for chimpanzees in eastern Africa. The shape of the mismatch distribution for chimpanzees in eastern

Pan troglodytes

Upper Guinea (including western Nigeria)



N = 78 (80)

Mean pairwise diff. = 12 (14)

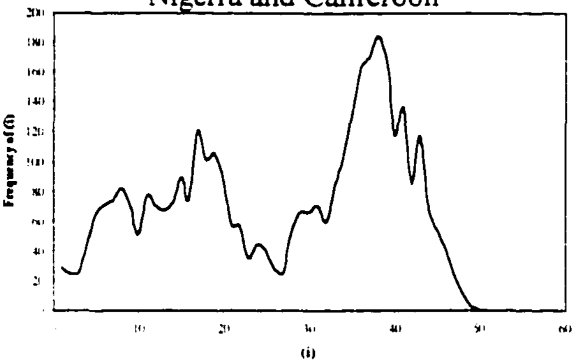
$\theta_0 = 4.93\ (5.808)$

$\tau = 7.603\ (8.806)$

$r = 0.004\ (0.002)$

$\chi^2 = 40446.078\ (48808.08)$, d.f. = 32 (39)

Nigeria and Cameroon



N = 86

Mean pairwise diff. = 27

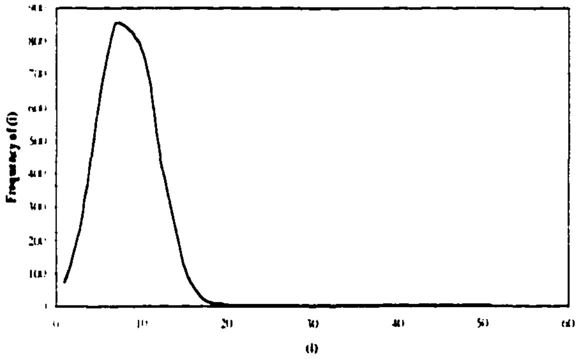
$\theta_0 = 11.855$

$\tau = 14.899$

$r = 0.002$

$\chi^2 = 12348.069$, d.f. = 66

Eastern Africa



N = 123

Mean pairwise diff. = 8

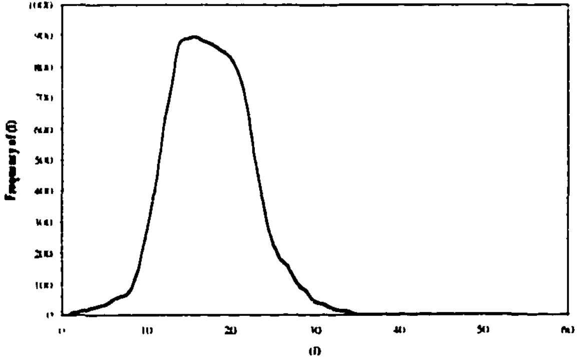
$\theta_0 = 1.473$

$\tau = 6.915$

$r = 0.004$

$\chi^2 = 315.992$, d.f. = 21

Homo sapiens



N = 135

Mean pairwise diff. = 18

$\theta_0 = 1.609$

$\tau = 8.343$

$r = 0.005$

Figure 4.8. Mismatch and intermatch distributions of chimpanzees in three geographical regions compared to humans. Human HV-I sequence data taken from Vigilant *et al.* (1991). Human mismatch distribution taken from Goldberg (1996). Numbers in parentheses for the Upper Guinea mismatch include western Nigeria.

Africa is very similar to the shape of human the mismatch distribution. The eastern African chimpanzee mismatch shows smooth unimodal shape and leftward position of the mismatch histogram, as well as low τ and r values. These properties of that mismatch and its similarities to the human mismatch distribution suggest a very rapid and recent expansion of chimpanzees throughout forests in eastern Africa (Goldberg 1996; Goldberg 1998).

The mismatch histogram in the upper panel of Figure 4.8 shows some evidence of bimodality among haplotypes of chimpanzees in Upper Guinea, but still retains a relatively smooth shape and low τ and r values. This histogram and its associated statistics indicate two properties about Upper Guinea chimpanzees. First, the slight bimodality of the mismatch suggests some population subdivision, multiple population expansions, multiple waves of migration from chimpanzees elsewhere, or some combination of the above (Rogers and Jorde 1995). Secondly, compared to the human and eastern African mismatch histograms, the higher τ values and more rightward position of the histogram suggest a more ancient population expansion of chimpanzees in Upper Guinea forests.

The dashed line in the upper panel of Figure 4.8 illustrates the mismatch distribution for chimpanzees in Upper Guinea and those in western Nigeria. The mismatch distribution for these combined samples has a similar shape to the mismatch limited to chimpanzees in Upper Guinea, but is located further to the right, and has slightly higher θ_0 and τ estimates. These mismatch distributions suggest that chimpanzees in western Nigeria may share a fairly ancient relationship with those in Upper Guinea. However, the small sample size of HV-I sequences from western Nigeria and their ambiguous relationships to other chimpanzees in western Africa prohibited further evaluation of these samples.

The second panel of Figure 4.8 illustrates the mismatch histogram for the entire Nigeria-Cameroon data set. This mismatch distribution is markedly different from the other mismatches shown in Figure 4.8. This histogram clearly shows a very ragged distribution, with two well-defined peaks that occur roughly at 20 and 37 pairwise site differences. At first glance, the bimodality of this histogram suggests that the Nigeria-Cameroon chimpanzee population contains two ancient and subdivided lineages. Therefore, this mismatch histogram falls under the category of an intermatch distribution (Marjoram and Donnelly 1994). The τ value corresponds to the leading wave in the intermatch distribution (Harpending *et al.* 1993; Rogers and Jorde 1995). Compared to the mismatch distributions generated for the Upper Guinea and eastern African chimpanzee populations, the τ value for the Nigeria-Cameroon intermatch distribution is over two times higher. Because τ is proportional to the time at which a population began to expand, the Nigeria-Cameroon population probably is at least twice as old as populations in Upper Guinea or those in eastern Africa.

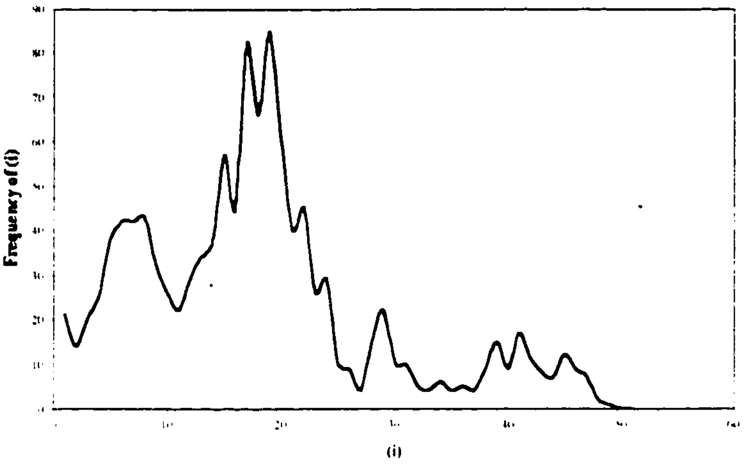
However, the very low *r* value for Nigeria-Cameroon mismatch is not concordant with that conclusion, and it is not clear why the *r* value is that low. One possibility is that the degree of subdivision is so high that the *r* statistic is inappropriate for very ancient lineages. A second possibility is that chimpanzee populations in Nigeria and Cameroon still show the effects of recent westerly and easterly migrations at the HV-I locus, and in fact, are less demographically stable than the Upper Guinea or eastern African regional populations.

Given that two lineages appear to be present in the Nigeria-Cameroon intermatch distribution shown in Figure 4.8, sequences of samples from across the region were divided along geographic boundaries. All of the previous analyses for the HV-I locus suggest a genetic break at the Sanaga River, and samples for the subdivided data set were separated along that boundary. Figure 4.9 shows mismatch distributions for samples from eastern Nigeria and western Cameroon north of the Sanaga River (top panel) and for samples from Cameroon south of the Sanaga River. Chi-square values for both mismatch distributions in Figure 4.9 suggest that modal HV-I sequence pairwise differences among chimpanzee populations in Nigeria and Cameroon differ from null expectations.

The top panel of Figure 4.9 shows the mismatch distribution for samples collected in eastern Nigeria and western Cameroon north of the Sanaga River. When these samples are separated from those in Cameroon south of the Sanaga River, the mismatch distribution is more similar to the mismatch distributions for sequences from chimpanzees in Upper Guinea and eastern Africa. The τ value is higher but comparable to other populations. The leading wave of the intermatch is still present (beginning at about 28 pairwise site difference), but to a lesser degree. While the histogram appears less smooth than histograms generated for chimpanzee populations in Upper Guinea and eastern Africa, the *r* value is equivalent.

The bottom panel of Figure 4.9 shows the mismatch distribution of sequences for samples collected in Cameroon south of the Sanaga River. There are three distinct peaks in the mismatch histogram. The leading two waves in this mismatch distribution probably represent the leading wave in the intermatch distribution shown in the second panel of Figure 4.8. The τ value for the southern Cameroon mismatch is nearly equal to that

Eastern Nigeria and western Cameroon north of the Sanaga River



N = 49

Mean pairwise diff. = 8

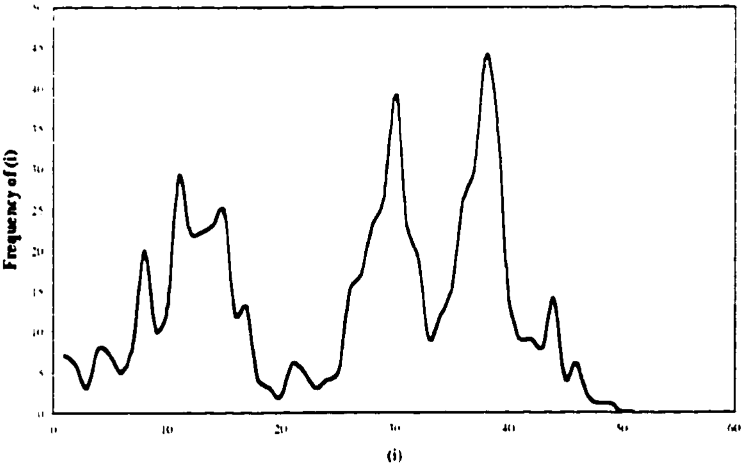
$\theta_0 = 9.762$

$\tau = 8.305$

$r = 0.004$

$\chi^2 = 958.546$, d.f. = 44

Cameroon South of the Sanaga River and east of the Mbaum River



N = 37

Mean pairwise diff. = 12

$\theta_0 = 11.216$

$\tau = 14.644$

$r = 0.006$

$\chi^2 = 1234.308$, d.f. = 42

Figure 4.9. Mismatch distributions for samples from Nigeria and Cameroon. Samples were divided into groups across the Sanaga River.

shown in the second panel of Figure 4.8. These observations imply that this chimpanzee population is much older than regional populations to the north, west and east. The higher *r* value (0.006) for sequences from this region suggest that the chimpanzee population in southern Cameroon is closer to reaching demographic equilibrium than in other regions. Compared to simulated *r* values, however, this population appears to have undergone a fairly recent population expansion.

Timing population divergence and expansion

Theory and methods

A variety of methods are available to estimate population divergence times and expansion dates using relative rates tests calibrated with universal and local molecular clocks (presented in Chapter 3) and coalescent models. However, determining accurate estimates is nearly impossible. Confidence limits are usually very large, and in some cases are contradictory. Many models assume random mating, equilibrium population size, exponentially distributed bifurcation times and a constant mutation rate (often estimated from between-lineage rates) (Hillis, Mable, and Moritz 1996). Mutation rates, especially at the HV-I locus, are difficult to determine with any precision. The HV-I locus evolves rapidly, and multiple substitutions between and within lineages can confound accurate rate estimates (Tamura and Nei 1993; Meyer, Weiss, and von Haesler 1999; Pesole *et al.* 1999). Consequently, estimates of population divergence dates and expansion times are best viewed as very rough approximations.

Errors surrounding the parameters of mismatch and intermatch distributions are large. However, it is possible to use these distributions to estimate population expansion dates. As noted previously in equations 4.8-4.10, the location and crest of the mismatch and intermatch waves are defined by τ, which is directly proportional to the time at which the population underwent expansion. Rogers and Harpending (1992) illustrated that τ can be used to approximate the time since population expansion by:

4.12) $$t = \tau/(2m_T\mu)$$

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

4.13) $$T_{(abs\ yrs)} = G(\tau/2m_T\mu)$$

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

Nucleotide mutation rates (μ) were taken from several sources (Vigilant *et al.* 1991; Ward *et al.* 1991; Goldberg 1996; Pesole *et al.* 1999). The estimates of Vigilant *et al.* (1991) and Ward *et al.* (1991) were derived from analogous regions of the human HV-I sequence and assume rate constancy between humans and chimpanzees. Vigilant *et al.* (1991) proposed a nucleotide divergence rate of 11.5% per million years, and a date of divergence between humans and chimpanzees of 6mya. Ward *et al.* (1991) propose a 33.5% nucleotide divergence rate, and a divergence between humans and chimpanzees of 4mya. Goldberg (1996) used a 22.5% median nucleotide divergence rate estimate to time the population expansion of eastern African chimpanzees. This estimate was adopted for this analysis to directly compare population expansion dates to that study. Finally, Pesole *et al.* (1999) analyzed mtDNA control region sequences of bonobos and chimpanzees and proposed a 3.48% divergence rate at the HV-I locus. They timed the separation of these species at 3.25mya, which is almost certainly too ancient. The assumed generation time for chimpanzees was 25 years.

Results and interpretation

Each of these mutation rates was used to calculate *T* from equation 4.13. Approximate population expansion dates are listed in Table 4.6. Population expansion dates varied widely depending on the mutation rate used in the calculations. However, these dates do suggest some interesting properties about the expansion of chimpanzees throughout Africa.

The combined HV-I sequences from Nigeria and Cameroon yield the most ancient population expansion date. The intermediate mutation rates (T_2 and T_3) suggest that populations from the entire Nigeria-Cameroon study area probably coalesced between 97kya and 187kya. When τ values for the subdivided data set were used to time population expansion, samples from Cameroon south of the Sanaga River and western equatorial Africa represent the most ancient lineage that may have expanded between 95kya and 183kya. Samples from eastern Nigeria and western Cameroon north of the Sanaga River represent a more youthful group that may have expanded between 54kya and 105kya. HV-I sequences of Upper Guinea chimpanzees imply an expansion of these chimpanzees between 76kya to 146kya (excluding samples from western Nigeria), and

Table 4.6. Approximate population expansion dates based on empirical estimates of τ.

	n	θ_0	τ	$T_{1(KYA)}$*	$T_{2(KYA)}$*	$T_{3(KYA)}$*	$T_{4(KYA)}$*
Pan troglodytes							
Upper Guinea†	78	4.93	7.603	51.43	75.78	146.30	487.67
Nigeria + Cameroon	86	11.855	14.899	65.43	96.95	187.23	624.00
Eastern Nigeria and Cameroon north of the Sanaga R.	49	9.762	8.305	36.58	54.20	104.67	349.00
Western Equatorial Africa including Cameroon south of the Sanaga R.	37	11.216	14.644	64.10	95.03	183.49	611.00
Eastern Africa	123	1.473	6.915	30.55	45.28	87.41	291.00
Homo sapiens	135	1.609	8.343	36.30	54.13	103.93	346.45

*Expansion dates are based on empirical estimates of τ. T_1 is based on Ward *et al.*'s (1991) nucleotide divergence rate of 33% per million years. T_2 used an intermediate nucleotide divergence rate of 22.25% per million years (Goldberg 1996). T_3 is based on Vilgilant *et al.*'s (1991) divergence rate of 11.5% per million years. T_4 is based on Pesole *et al.*'s (1999) divergence rate of 3.48% per million years. All dates are given in thousands of years.

†Samples from western Nigeria were excluded from these calculations due to their ambiguous relationship with Upper Guinea chimpanzees.

represent a group that is the same age or slightly older than chimpanzees further east in Nigeria and western Cameroon. This observation supports Gagneux's (1998) hypothesis that chimpanzees in Upper Guinea may represent a moderately ancient lineage that has experienced multiple waves of migration from populations further to the east and/or has experienced smaller independent population expansions over a long time period. Finally, chimpanzee HV-I sequences from eastern Africa represent the most youthful group with an expansion date probably between 45kya and 87kya. These dates coincide with the population expansion dates given by Goldberg (1996).

These estimates are consistent with order of the relative divergence dates presented in Table 3.2. Estimates of population expansion dates using the parameter, τ, are based on the position and size of the crest of the wave in a mismatch distribution (Rogers and Jorde 1995). In contrast, estimates of relative divergence dates are calculated using maximal or mean genetic differences calculated from the most basal nodes of a gene tree (Ruvolo *et al.* 1993). Consequently, the relative divergence dates presented in Chapter 3 estimate a date of the origin of the inferred ancestral haplotype for each putative lineage, and therefore, correspond to the left tail of the mismatch distributions (T. Disotell, pers. comm.). The differences between these two methods result in smaller population expansion estimates based on mismatch distributions. At the HV-I locus in humans, for example, the relative divergence date estimate implies a human mitochondrial ancestor roughly 298kya (Ruvolo et al. 1993). In contrast, the mismatch distribution for the human HV-I locus implies a human population expansion between 30kya and 150kya (Rogers and Jorde 1995), or between 51kya and 103kya (Table 4.6).

Despite the fact that there are numerical differences between the relative divergence dates and the population expansion dates calculated from the mismatch and intermatch distributions, both methods suggest that the most ancient population of chimpanzees is located in western equatorial Africa. Chimpanzee populations in Upper Guinea and in eastern Nigeria and western Cameroon represent the next most ancient groups. Populations in eastern Africa represent the most youthful group.

Conclusions

The population genetic analyses presented in this chapter largely support the phylogenetic reconstructions presented in Chapter 3. The apportionment of genetic diversity at the HV-I locus suggests that two deeply divergent chimpanzee lineages are present across Africa: a western African lineage and a central African lineage. The ranges of these lineages converge in central Cameroon. Both the AMOVA and spatial autocorrelation analyses suggest that the Sanaga River is the most likely geographical barrier that has influenced the distributions of these lineages. In fact, over 30% of the genetic diversity at the HV-I locus detected by AMOVA is attributable to differences between populations separated by the Sanaga River.

The Sanaga River has not completely inhibited gene flow between the western and central African lineages, but migration across the Sanaga probably has been rare. Populations on either side of the Sanaga exchange only a single migrant, on average, per generation. The phylogenetic reconstructions presented in Chapter 3 suggest that migration has occurred more frequently at Manb'ra (map location, 13). There is also some evidence of gene flow between the western and central African lineages across large distances at the Cross River National Park in southeastern Nigeria (map locations, 3 and 4) and at the Dja Biosphere Reserve in southern Cameroon (map location, 16). Despite this evidence of gene flow, the Sanaga River probably has had a long term influence on chimpanzee distribution patterns. The population expansion dates calculated from the mismatch distributions imply that the Sanaga River probably has affected the distributions of the western and central African chimpanzee lineages for the last 100 to 200ky, but it may have had a longer influence on chimpanzee distribution patterns.

The influence of the Niger River and of the Dahomey Gap on chimpanzee distribution patterns could not be fully evaluated in this study due the small sample size of chimpanzee HV-I sequences from western Nigeria. However, there appears to be more genetic differences between chimpanzee populations in Upper Guinea and those in eastern Nigeria and western Cameroon than between populations in western equatorial and eastern Africa. The linearized F_{ST} D values illustrate that chimpanzee populations in western Africa coalesce earlier than those western equatorial and eastern Africa. In addition, no HV-I sequences were shared between chimpanzees in Upper Guinea and those in Nigeria or Cameroon. Yet, shared haplotypes were detected between populations separated by over 1,000 km in Upper Guinea forests. Shared haplotypes were only detected over a 200 km span of the Cameroon Highlands in Nigeria. It seems possible that these chimpanzees are not part of the same population, or perhaps even of the same lineage.

In contrast, the pairwise linearized F_{ST} D values calculated for the western equatorial and eastern African regional chimpanzee populations suggest that these two groups share a more recent relationship than populations further to the west. In addition, eurytopic haplotypes were found in populations in eastern Africa separated by nearly 600 km. The presence of two "eastern African-like" HV-I sequences from central Cameroon in the phylogenetic reconstructions presented in Chapter 3 provide further support for the hypothesis that western equatorial and eastern African chimpanzees share a fairly recent relationship.

The mismatch and intermatch distributions suggest several interesting properties of recent chimpanzee evolution. HV-I sequences from western equatorial Africa are more diverse than in any other region, and imply that chimpanzees from this area possess the most ancient HV-I haplotypes of all modern chimpanzee populations. These observations combined with the largest population expansion date imply that chimpanzees in western equatorial Africa represent the most ancient of all chimpanzee populations. In addition, the multimodality of the mismatch distribution for samples from this area suggest that chimpanzees in western equatorial Africa have had a more stable paleodemographic history than other populations. In contrast, the mismatch distributions and population expansion date estimates suggest that chimpanzees in eastern Africa represent the most recent radiation of this species into African forests.

Chimpanzees in Upper Guinea and those in Nigeria and western Cameroon probably have had a more complex and ancient relationship. The mismatch distributions suggest that chimpanzees in eastern Nigeria and western Cameroon represent a relatively ancient radiation of chimpanzees. Chimpanzees probably colonized forests in western Africa sometime after the colonization of forests in western equatorial Africa. However, the dissimilar shapes of the mismatch distributions for chimpanzees limited to Upper Guinea and those in eastern Nigeria and western Cameroon suggest that populations from these areas have experienced markedly different paleodemographic histories. The slight bimodality of the mismatch distribution of chimpanzees from Upper Guinea suggests that this group may have undergone a larger more recent population expansion following their initial radiation into western African forests. The multimodality of the mismatch distribution of chimpanzees in eastern Nigeria and western Cameroon suggest that these populations have been more demographically stable than those further to the west following their initial expansion. These observations provide compelling, but not conclusive, evidence that a further division of the western African lineage into two separate lineages that are separated by the lower Niger River in Nigeria (or by the Dahomey Gap) may be warranted.

CHAPTER 5

Nuclear DNA Diversity among Chimpanzees in Western Africa

Introduction

Mitochondrial DNA sequences produced during this study represent only a small portion of a single genome. A complete analysis of a species' genetic diversity should include multiple genetic systems, preferably including loci that are independent, and that have different mutation rates (Avise 1994, 1998; Hoelzer, Wallman, and Melnick 1998). In order to achieve that goal, this chapter presents analyses of several rapidly evolving microsatellite, or STR, loci that occur ubiquitously throughout animal genomes (Jarne and Lagoda 1996; Goldstein and Pollock 1997). In order to standardize comparisons of genetic diversity between the HV-I and STR databases. STR loci were examined whenever possible from samples of chimpanzees from Nigeria and Cameroon that also produced reliable mtDNA HV-I sequences. One hundred and fifteen individual samples were genotyped for up to 10 different STR loci that are listed in Table 2.5. Appendix 4 lists the genotypes of all samples. Samples of chimpanzees in Upper Guinea (n = 89) and from Gombe (n = 29) were obtained from Gagneux (1998) and Morin (1992).

For this study, ten heterologous human STR markers were used to examine chimpanzee genetic diversity. It is possible to use such STR loci to make inferences about the evolutionary history of closely related lineages (e.g., Paszek *et al.* 1999; Goldstein *et al.* 1999). However, the accuracy of those inferences relies on a broad understanding of the mutational processes of each locus (Weber and Wong 1993; Takezaki and Nei 1996; Goldstein and Pollock 1997; Paszek *et al.* 1999). If lineages of interest diverged distantly in the past, STR markers may be of little value in accurately reconstructing phylogenetic relationships due to size constraints upon STR loci. Specifically, STR markers may be of little use in these cases because shared allele sizes are likely to be the result of homoplasy rather than homology (Bruford and Wayne 1993; Estoup *et al.* 1995; Garza, Slatkin, and Freimer 1995; Rubinsztein *et al.* 1995; Garza and Freimer 1996; Crawford *et al.* 1998; Pritchard *et al.* 1999; Taylor, Sanny, and Breden 1999). For these reasons, phylogenetic reconstructions of the STR database are not presented in this chapter. Instead, I present analyses that examine the apportionment of genetic diversity at ten STR loci and assess the degree to which potential biogeographic boundaries may have influenced the distribution of chimpanzees in Nigeria and Cameroon. Although several different types of analyses are presented in this chapter, most of these analyses rely on the calculation of fixation indices (F-statistics).

Summaries of genetic diversity

Theory and methods

There are several methods available to examine the apportionment of genetic diversity at STR loci. One of the most basic measures examines the distribution of allele frequencies at different loci within and between populations. Observed heterozygosity, expected heterozygosity and deviations from Hardy-Weinberg equilibrium expectations can be calculated from allele frequency summary data (Smith 1989; Avise 1994; Weir 1996a, 1996b).

There are several ways to calculate heterozygosity. Within the Arlequin program (Schneider *et al.* 1997), heterozygosity can be calculated by the number of different alleles or the number of pairwise differences between individuals. In both cases, heterozygosity is defined as:

5.1) $$d_{xy} = \sum_{i=1}^{L} \sigma_{xy}(i)$$

where $\sigma_{xy}(i)$ is the Kronecker function. This function is equal to one if the alleles of the i-th locus are identical for both genotypes, and equal to zero otherwise. When this heterozygosity estimate is incorporated into a genetic structure analysis in a multilocus data set, it results in estimating weighted F_{ST} statistics over all loci (Weir and Cockerham 1984; Michalakis and Excoffier 1996).

Hardy-Weinberg equilibrium (HWE) refers to a stability of genotypes achieved in populations of infinite size after one generation of panmictic mating (Hillis, Moritz, and Mable 1996). Deviation from HWE suggests a population or a group of populations whose genetic structure may be influenced by microevolutionary processes such as immigration, emigration, selection or mutation (Smith 1989; Weir 1996a). There are several methods available to evaluate deviations from HWE. Traditional HWE models calculate empirical deviations from Hardy-Weinberg expectations relative to HWE values for a theoretical population at equilibrium (reviewed in Weir 1996). However, comparing empirical HWE values to a null distribution of HWE values calculated from randomly generated allele frequency distributions yields more accurate estimates (Guo and Thompson 1992; Weir 1996).

Guo and Thompson (1992) proposed using a maximum likelihood Markov chain method to estimate deviations from HWE. In this model, a two-by-two contingency table is built during the analysis, but is extended to a triangular contingency table of arbitrary size. Null contingency tables are constructed from the original data set by selecting two rows and two columns at random. However, instead of enumerating all possible contingency tables, a Markov chain is used to efficiently explore the space of all possible tables. This exploration results in a random "walk" through the space of the contingency tables.

Within the Arlequin computer program, the random walk of the Markov chain is executed such that the probability of a visit to a particular table corresponds to its actual probability under the null hypothesis of HWE (Schneider *et al.* 1997). The significance of the deviation of a population from HWE is calculated as the proportion of visited tables having a probability smaller or equal to the observed contingency table (Guo and Thompson 1992). In this model, the probability of obtaining the observed table under the null-hypothesis of HWE is defined as:

5.2) $$Lo = \frac{n! \prod_{i=1}^{k} n_i*!}{\prod_{i=1}^{k} \prod_{j=1}^{i} n_{ij}!}(2H)$$

where H is the number of heterozygous individuals.

Results and interpretation

Several descriptive statistics were calculated for the STR database using the computer program Arlequin (Schneider *et al.* 1997). Genetic diversity, observed heterozygosity, expected heterozygosity and deviations from Hardy-Weinberg null expectations within each population were calculated based on the mean number of different alleles per locus (equation 5.1). Appendix 5 summarizes allele size variation, provides an estimate of heterozygosity and reports the probability of deviation from HWE using the Guo and Thompson (1992) Markov chain method for each locus and each population. Table 5.1 provides a summary of the heterozygosity estimates and HWE calculations given in Appendix 5. Figures 5.1-5.3 present allele frequency distributions for the 10 STR loci. The histograms presented in these figures show allele frequencies for samples collected in eastern Nigeria and western Cameroon north of the Sanaga River (open bars) and those collected in Cameroon south of the Sanaga River (closed bars). Samples from western Nigeria were grouped into a single population (striped bars).

Table 5.1 and Figures 5.1-5.3 reveal a high-level of genetic diversity among chimpanzees in Nigeria and Cameroon. The 10 STR loci have an average of 8 alleles, with a range of 5 to 15 alleles per locus. With few exceptions, alleles for a given locus are not limited to any single population or group of populations. Heterozygosity estimates range widely from 17 – 96%, with an average heterozygosity estimate of 74%. In most cases, observed heterozygosity estimates do not differ significantly from expected levels of heterozygosity or from Hardy-Weinberg null expectations. There is

Table 5.1. STR allelic diversity, heterozygosity estimates and deviations from Hardy-Weinberg proportions

Map locations given in parentheses are listed in Table 2.1 and Figure 2.2.

Sample Location	APOA2	D4S1652	D7S1809	D9S303	D11S1984	HUMPLA2A	D13S317	D16S265	D16S539	D20S470
Western Nigeria (1+)*										
Alleles	6	5	4	2	3	3	No data	7	3	5
Genes	12	10	10	8	6	6		8	10	6
Heterozygosity	0.85	0.87	0.78	0.57	0.80	0.73		0.96	0.51	0.93
Exact HWE p	0.39	0.06	0.01	0.08	0.07	0.19		1.00	0.33	1.00
Gashaka-Gumti (7-9)										
Alleles	6	9	5	5	5	4	7	4	3	6
Genes	18	14	10	8	10	12	16	10	18	10
Heterozygosity	0.80	0.93	0.82	0.86	0.67	0.74	0.90	0.78	0.52	0.89
Exact HWE p	0.05	0.52	0.35	0.65	0.63	0.27	0.00	1.00	1.00	0.13
Ngel Nyaki (6)										
Alleles	6	7	No data	4	No data	2	No data	3	4	4
Genes	10	8		4		4		4	8	4
Heterozygosity	0.88	0.96		1.00		0.50		0.83	0.64	1.00
Exact HWE p	0.63	0.17		1.00		1.00		1.00	0.44	1.00
Akoh Zanto (5)										
Alleles	4	9	5	7	6	3	4	8	4	12
Genes	20	20	12	14	14	16	12	20	22	16
Heterozygosity	0.75	0.89	0.80	0.91	0.87	0.43	0.77	0.79	0.71	0.96
Exact HWE p	0.42	0.01	0.02	0.00	0.00	0.01	0.20	0.01	0.50	0.36
Cross River (3-4)										
Alleles	5	6	2	5	No data	2	5	No data	4	5
Genes	20	20	12	10		6	8		20	14
Heterozygosity	0.65	0.84	0.17	0.87		0.53	0.89		0.54	0.73
Exact HWE p	0.49	0.19	1.00	0.00		0.20	0.01		0.14	0.01
Banyang Mbo (10)										
Alleles	2	3	3	No data		5	3	No data	4	5
Genes	10	6	4			10	4		8	8
Heterozygosity	0.36	0.80	0.83			0.84	0.83		0.64	0.86
Exact HWE p	1.00	0.46	0.34			0.35	0.33		0.43	0.08

Table 5.1. STR allelic diversity, heterozygosity estimates and deviations from Hardy-Weinberg proportions (continued)

Sample Location	APOA2	D4S1652	D7S1809	D9S303	D11S1984	HUMPLA2A	D13S317	D16S265	D16S539	D20S470
Mosse (12)										
Alleles	7	No data	5	No data	No data	5	6	7	5	7
Genes	12		10			16	8	16	18	12
Heterozygosity	0.91		0.82			0.53	0.93	0.83	0.72	0.93
Exact HWE p	1.00		0.07			0.01	0.02	0.25	0.03	0.14
Mt. Cameroon (11)										
Alleles	6	No data	5	No data	No data	6	6	5	5	8
Genes	7		8			18	14	18	22	14
Heterozygosity	0.93		0.86			0.75	0.85	0.78	0.70	0.87
Exact HWE p	0.05		0.84			0.83	0.09	0.63	0.83	0.87
Manb'ra (13)										
Alleles	6	3	5	5	No data	3	9	6	5	11
Genes	18	10	8	10		14	14	18	28	22
Heterozygosity	0.83	0.60	0.86	0.84		0.67	0.93	0.81	0.72	0.88
Exact HWE p	0.12	1.00	0.09	0.36		0.01	0.00	0.30	0.03	0.64
DEFR (14)										
Alleles	6	5	4	8	6	4	7	3	3	8
Genes	18	8	12	14	10	20	10	14	18	16
Heterozygosity	0.77	0.86	0.68	0.90	0.89	0.50	0.91	0.69	0.50	0.88
Exact HWE p	0.97	0.09	0.01	0.21	0.01	0.03	0.04	0.48	0.63	0.10
Campo (14)										
Alleles	6	10	4	5	4	3	4	5	4	8
Genes	22	18	8	12	8	10	12	6	20	14
Heterozygosity	0.79	0.92	0.77	0.72	0.82	0.67	0.77	0.93	0.60	0.92
Exact HWE p	0.00	0.00	0.78	1.00	0.03	0.05	0.09	0.21	0.05	0.00
Dja (16)										
Alleles	5	3	3	8	3	3	4	5	4	9
Genes	18	6	6	14	10	8	6	14	22	20
Heterozygosity	0.78	0.73	0.73	0.82	0.71	0.71	0.87	0.73	0.46	0.92
Exact HWE p	0.27	0.20	1	0.00	0.01	0.03	0.07	0.33	0.40	0.01

*Western Nigeria includes samples from Ise Forest Reserve (map location, 1) and preserved skins obtained in western Nigeria.

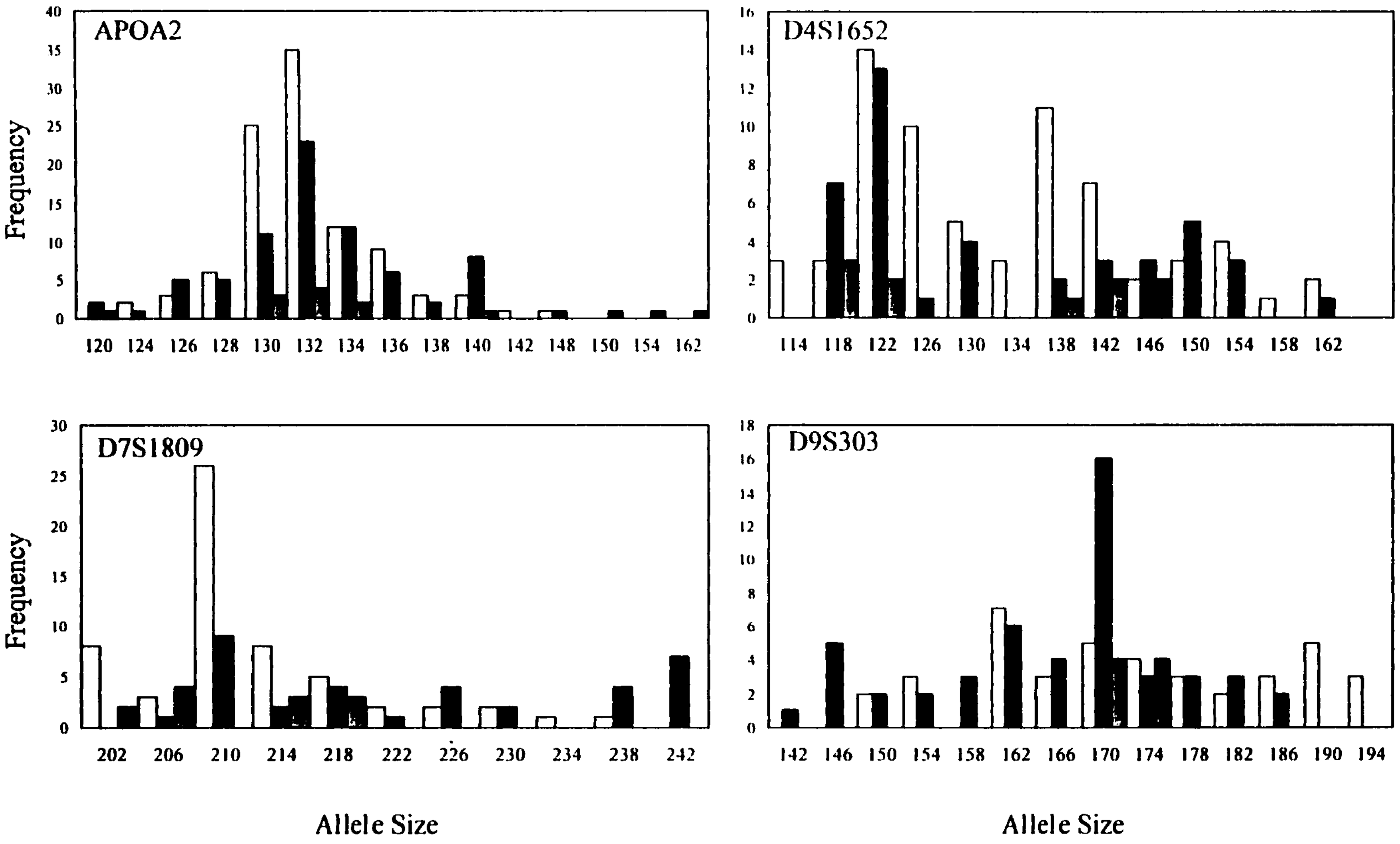


Figure 5.1. Allele frequency distributions for the APOA2, D4S1652, D7S1809 and D9S303 loci. Striped bars represent samples from western Nigeria. Open bars represent samples from eastern Nigeria and western Cameroon north of the Sanaga River. Black bars represent samples from Cameroon south and east of the Sanaga River.

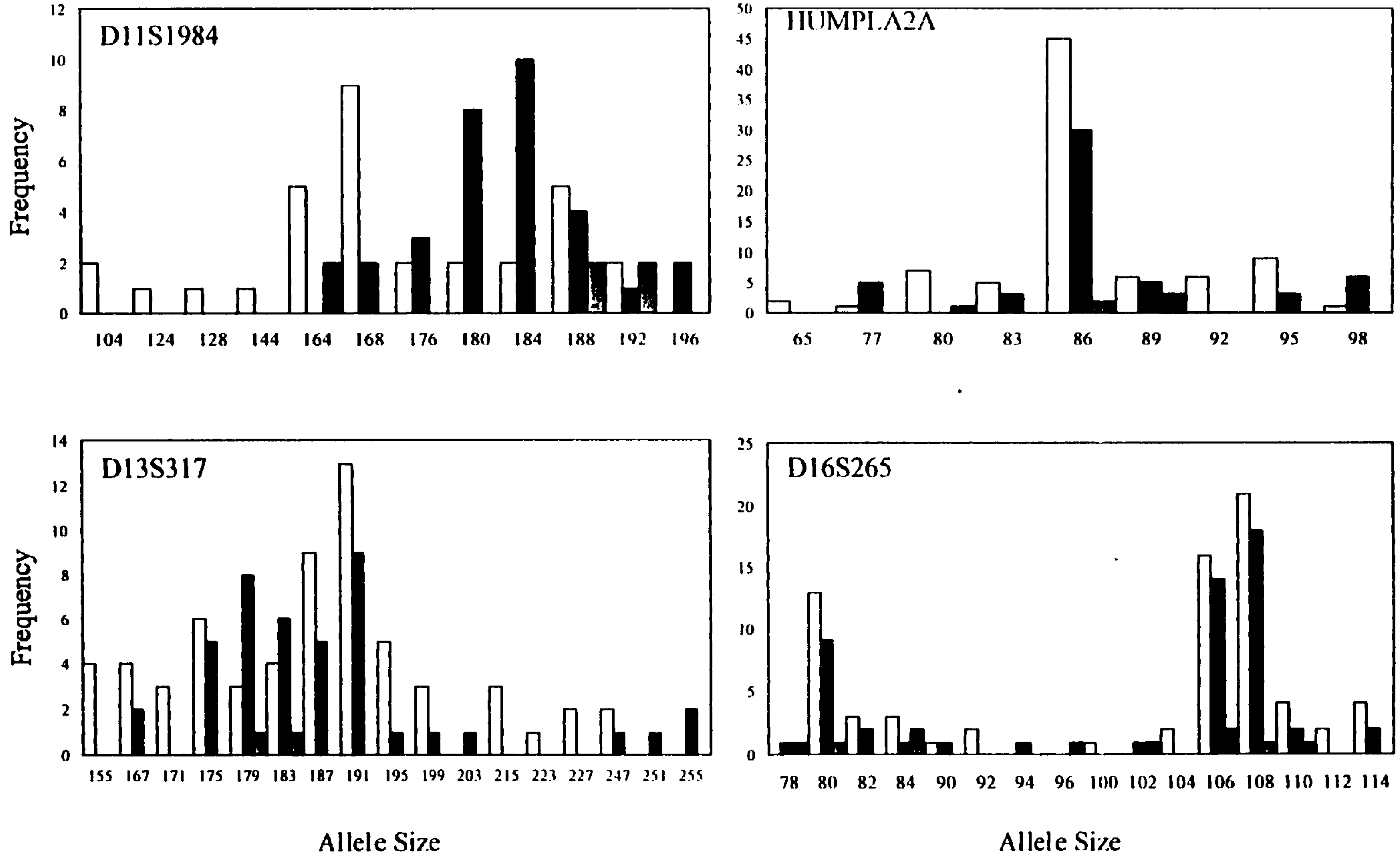


Figure 5.2. Allele frequency distributions for the D1S11984, HUMPLA2A, D13S317 and D16S265 loci. Striped bars represent samples from western Nigeria. Open bars represent samples from eastern Nigeria and western Cameroon north of the Sanaga River. Black bars represent samples from Cameroon south and east of the Sanaga River.

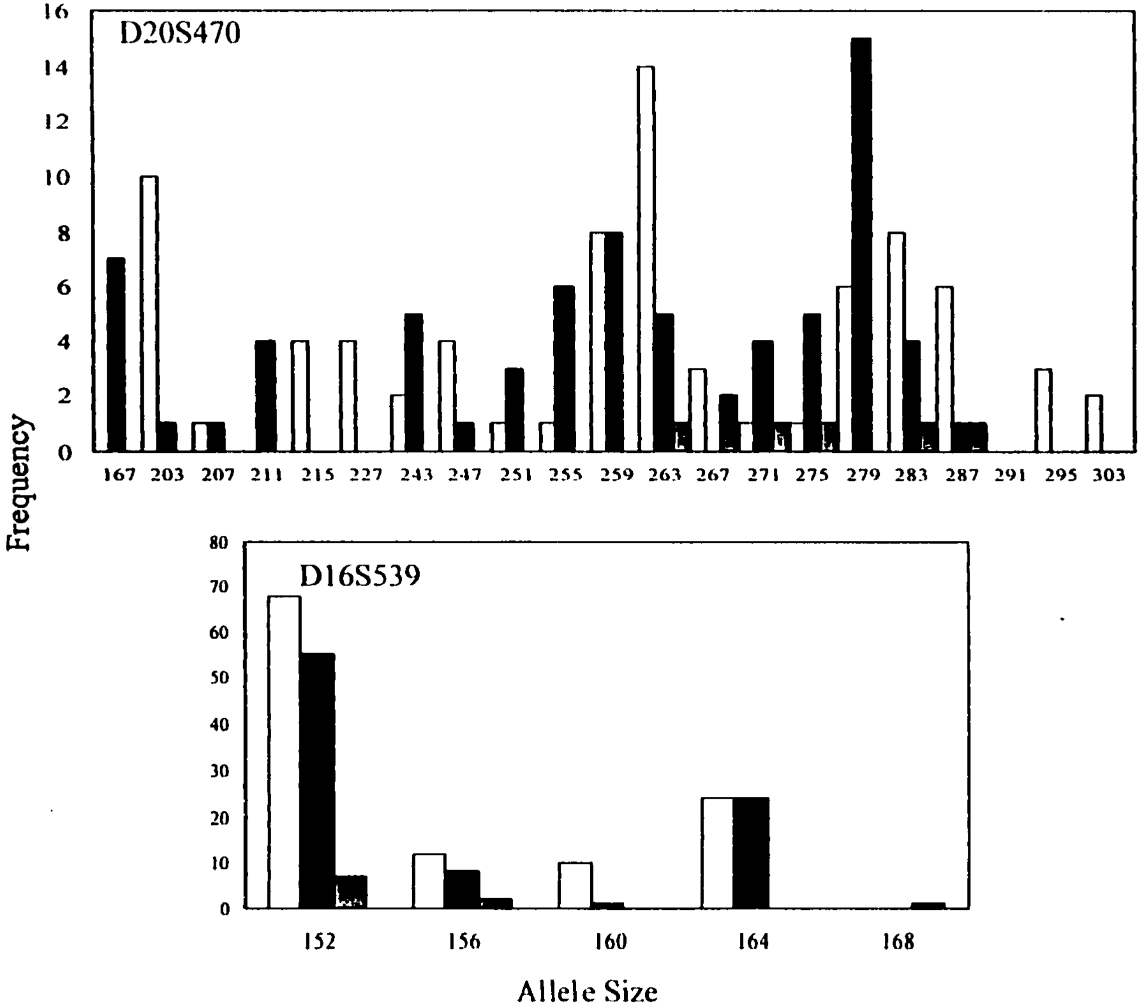


Figure 5.3. Allele frequency distributions for the D16S539 and D20S470 loci. Striped bars represent samples from western Nigeria. Open bars represent samples from eastern Nigeria and western Cameroon north of the Sanaga River. Black bars represent samples from Cameroon south and east of the Sanaga River.

also no detectable geographical pattern or trend for any locus for those populations that do deviate from expected Hardy-Weinberg proportions.

Population pairwise estimates of genetic differentiation

Theory and methods

F-Statistics

Estimates of genetic diversity are useful for identifying unique alleles, loci, and populations that may be influenced by microevolutionary processes. However, these statistics provide little insight about the relationships between different populations. Wright (1951) developed a measure of genetic differentiation based on an island model of populations that estimates the extent to which allele frequencies deviate from Hardy-Weinberg equilibrium expectations. This model is based on the idea that small, isolated populations tend to become differentiated from one another due to the random loss of alleles. In order to measure the degree of population subdivision, Wright (1951, 1969) proposed using three fixation indices, F_{IT}, F_{IS} and F_{ST} (see Chapter 3). F_{ST} is estimated by the ratio of heterozygosity calculated at different levels of population substructure, and defined as:

5.3) $$F_{ST} = \frac{H_T - H_S}{H_T}$$

where H_T is the heterozygosity in the total population and H_S is the average heterozygosity over all subpopulations. Wright's fixation indices assume an island population model, infinite alleles, subpopulations of identical size, equal levels of migration between subpopulations, and that all subpopulations are at equilibrium with respect to genetic drift and migration (Smith 1989; Weir 1996b).

Wright's (1951, 1969) traditional model has been adapted to a wide range of genetic systems, although the mathematical notation accompanying these models varies. For convenience, these models generally are referred to as F-statistics (Avise 1994; Weir 1996a, 1996b). These models tend to produce robust results under different evolutionary assumptions, with finite sample sizes and with multiple loci, even when model assumptions have been violated (Cockerham 1969; Nei 1977; Nei 1978; Weir and Cockerham 1984; Nei 1987; Lynch and Crease 1990; Slatkin 1995; Goodman 1997).

For this project, fixation indices were calculated using an analogous model derived from Cockerham (1969). He proposed three fixation indices θ, F and f put into the context of an analysis of variance. In his model, θ is analogous to F_{ST}, F is analogous to F_{IT}, and f is analogous to F_{IS}. For large numbers of samples approximate expressions of F-statistics for each allele can be calculated by:

5.4) $$\theta = \frac{s_A^2}{\bar{p}_A(1-\bar{p}_A)}$$

5.5) $$F = \frac{\bar{H}_A}{\bar{p}_A(1-\bar{p}_A)}$$

5.6) $$f = \frac{F-\theta}{1-\theta}$$

where, s_A^2 represents the variance of allelic frequencies over populations, $\bar{p}_A$ represents sample allelic frequencies in a sample of n_i from the ith population, and H_A represents the frequency of heterozygous individuals that have allele A. Similarly, these equations can be expressed in an analysis of variance format by:

5.7) $$\sigma_c^2 = (1-F)\sigma^2$$

5.8) $$\sigma_b^2 = (F-\theta)\sigma^2$$

5.9) $$\sigma_a^2 = \theta\sigma^2$$

where σ_a^2, σ_b^2, and σ_c^2 represent the variances between populations among groups, among individuals between populations but within groups and among individuals within populations, respectively, and $\sigma^2 = \sigma_a^2 + \sigma_b^2 + \sigma_c^2$. Finally, Cockerham's fixation indices can be expressed as variance terms by:

5.10) $$\theta = \frac{\sigma_a^2}{\sigma^2}$$

5.11) $$F = \frac{\sigma_a^2 + \sigma_b^2}{\sigma^2}$$

5.12) $$f = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2}$$

Cockerham (1969) reviewed the statistical and biological meaning of different fixation-index values. Departures from Hardy-Weinberg expectations are indicated by positive fixation indices; and suggest that microevolutionary processes may be acting to produce population substructure. For example, positive θ/F_{ST} values may be the result of genetic drift between populations, while positive f/F_{IS} values may result from inbreeding and a lack of immigration (Weir 1996a, 1996b).

In Wright's traditional model, F_{ST} must always be greater than or equal to zero (Cockerham 1969; Weir and Cockerham 1984; Weir 1996a). It is possible, however, to obtain negative θ/F_{ST} values. Negative variance components can sometimes occur, because they are actually covariances. Their associated fixation indices may be viewed as correlation coefficients (Weir 1996a, 1996b; L. Excoffier pers. comm.; P. Smouse pers. comm.). Negative fixation indices may be present within a pairwise distance matrix when sample sizes are limited. In that case, the negative values may reflect variances within subpopulations that are larger than the total population variance (Cockerham 1969; Weir and Cockerham 1984; Nei 1987; Weir 1996a). Negative fixation indices also can have biological meaning. Because the true value of the test statistic is zero, negative variance components can occur in the absence of genetic structure (Weir 1996a, 1996b; L. Excoffier pers. comm.; P. Smouse pers. comm.). Negative fixation indices also may indicate that more alleles are shared between rather than within populations (Weir 1996b). This happens commonly in out-crossing organisms where migration patterns violate the assumption of some models that populations have remained isolated following the split of the ancestral population into daughter populations (Cockerham 1969; Weir and Cockerham 1984; Nei 1987; Weir 1996b). Negative fixation indices are possible for all hierarchical analyses of genetic differentiation presented in this chapter because all the variance components are actually covariances, and the fixation indices are correlation coefficients (Schneider *et al.* 1997; L. Excoffier pers. comm.).

Slatkin's R_{ST}

One of the key assumptions of Wright's model is that mutations accumulate symmetrically at very low rate, and that different alleles are equidistant (Weir 1996a, 1996b). Moreover, this model assumes that the mutation process erases any memory of the prior allelic state, such that excess genetic similarity between populations, as measured by F_{ST}, can be attributed to migration or to historical association (Slatkin 1994a, 1995). However, the mutational process at many STR loci does not appear to conform to the classical model. STR loci mutate rapidly. Mutation rates at STR loci have been calculated to vary between 10^{-2} to 10^{-5} mutations per generation, depending on the repeat motif and the evolutionary constraints on the locus (Chakraborty *et al.* 1997; Schug *et al.* 1998).

In addition, a growing body of evidence suggests that the size of a new mutant allele depends upon the size of the allele that mutated. Because STR loci occur in a tandemly-repeated array, differences in allele sizes may be attributable to differences in repeat numbers. Consequently, STR loci are believed to mutate in a step-wise mutational model where repeat units are lost or added during the mutational process (Queller, Strassman, and Hughes 1993; Valdes, Slatkin, and Freimer 1993; Goldstein *et al.* 1995; Jarne and Lagoda 1996; Bossart and Powell 1998). However, there is considerable evidence to suggest that STR loci do not always conform to the stepwise mutation model, and when heterologous markers are used, violations of the stepwise mutation model may be very common (Amos and Rubinstzein 1996; Amos *et al.* 1996; Goldstein and Pollock 1997; Zhivotovsky, Feldman, and Grishechkin 1997; Brinkmann *et al.* 1998; Nielsen and Palsboll 1999).

The stepwise mutation model is rooted in the observation that there appears to be a positive asymmetry in the STR mutation process (Rubinsztein *et al.* 1995; Goldstein and Pollock 1997). In other words, when mutations occur at STR loci there is a trend to add one or two repeats. Over time, positive asymmetry leads to an increase in allele size. Differences in allele sizes may, therefore, be proportional to the time at which subpopulations coalescence.

Chapter 4 detailed Slatkin's (1991) linearized F_{ST} model for haploid data. This model assumes that mutations accumulate in a clock-like fashion, such that the number of site differences in a sample is proportional to the time at which the populations diverged. In this model, F_{ST} values are expressed as differences in the coalescent times of different populations. Given that STR allele size variation may be proportional to the time at which two populations diverged, Slatkin (1995) proposed a measure of population subdivision, R_{ST}, that incorporates divergence time between allele sizes with the stepwise mutation model. R_{ST} is analogous to Wright's F_{ST} model and capitalizes on the properties of the linearized F_{ST} model. R_{ST} is defined as:

5.13) $$R_{ST} = \frac{\bar{S} - S_W}{S_W}$$

where S_W and $\bar{S}$ are the average squared difference in allele size between pairs of genes within populations and between pairs of genes taken from a collection of P populations, respectively. As an analogous measure of R$_{ST}$, Michalakis and Excoffier (1996) proposed using the sum of the squared size difference between allele sizes. Their analogous measure is defined as:

5.14) $$d_{xy} = \sum_{i=1}^{L} (a_{xi} - a_{yi})^2$$

where a_{xi} is the number of repeats of the allele for the i-th locus.

R$_{ST}$ is difficult to apply to multilocus genotypes (Slatkin 1995). The main difficulty with this distance measure is its high variance, partly due to its dependence on variation within populations. In addition, because population sizes are likely to vary, the inclusion of an intrapopulation variance term can obscure the relationship between coalescent time and the observed differentiation value (Goldstein and Pollock 1997). Both Mickalakis and Excoffier (1996) and Goodman (1997) have standardized the R$_{ST}$ model for use with multilocus genotypes. Mickalakis and Exoffier (1996) followed Slatkin (1995) and proposed using a weighted average across all loci from Equation 5.12 when calculating R$_{ST}$ for multilocus genotypes. Goodman (1997) proposed using allele size as a fraction of the standard deviation for each locus. Both methods reduce the possibility of differences in the variance between loci or of differences in population size obscuring the variance that is attributable to population subdivision.

Results and interpretation

To test for interpopulational differentiation, both the F_{ST} and R_{ST} models were applied to genotypes from samples collected in Nigeria and Cameroon. All calculations were performed with the computer program Arlequin (Schneider *et al.* 1997) and were confirmed by the program MICROSAT (Minch *et al.* 1996). All F_{ST} and R_{ST} values were tested for statistical significance against 16,000 random permutations of the data using the Arlequin program. Table 5.2 presents a matrix of pairwise F_{ST} values based on allele frequencies and a matrix of pairwise R_{ST} values based on the sum of the squared difference between allele sizes. This table includes F_{ST} and R_{ST} statistics for the full multilocus database.

Pairwise comparisons of sampling locations suggest high levels of genetic diversity among chimpanzees in Nigeria and Cameroon, but very little genetic differentiation. F_{ST} estimates range from –0.376 to 0.075, with a pooled estimate of –0.075. Similarly, R_{ST} estimates reveal little population subdivision in pairwise comparisons. R_{ST} values range from –0.459 to 0.233, with a pooled estimate of 0.094. Both of the pooled estimates are not significantly different from the null distribution of differentiation values. For those comparisons that are statistically significant in both the F_{ST} and R_{ST} models, there is no detectable pattern related to geographic distance or to obvious population isolation. In the R_{ST} matrix, 33% of the significantly different pairwise comparisons are those that included genotypes obtained from samples collected in Cross River National Park in southeastern Nigeria (CRNP, Figure 2.2, map locations 3 and 4). All samples from Cross River National Park were checked for accuracy by multiple amplifications of each STR locus via PCR.

Table 5.2. Pairwise comparisons of genetic subdivision for 10 STR loci

Population Pairwise F_{ST} values*

	ise†	ggnp†	ngny†	akzn†	crnp†	bym†	mosse†	mtcm†	man†	defr†	campo†
ggnp	-0.047										
ngny	-0.046	-0.175									
akzn	**0.038**	**0.038**	-0.174								
crnp	0.024	-0.024	0.027	-0.079							
bym	-0.076	**0.051**	-0.080	0.001	-0.153						
mosse	-0.141	-0.118	-0.347	-0.166	-0.278	0.050					
mtcm	-0.265	-0.111	-0.436	-0.159	-0.376	0.020	0.007				
man	-0.097	-0.038	-0.110	-0.082	-0.070	0.036	-0.024	-0.052			
defr	**0.051**	**0.030**	-0.181	0.002	-0.139	0.007	-0.093	-0.117	-0.052		
campo	0.000	**0.028**	-0.039	-0.004	**0.075**	0.026	-0.164	-0.233	-0.049	-0.002	
dbr	-0.068	-0.057	-0.163	-0.025	-0.079	-0.034	-0.059	-0.157	0.000	-0.001	-0.037

Population Pairwise Standardized R_{ST} values*

	ise	ggnp	ngny	akzn	crnp	bym	mosse	mtcm	man	defr	campo
ggnp	-0.013										
ngny	**0.203**	-0.459									
akzn	-0.052	0.036	-0.380								
crnp	**0.183**	**0.223**	-0.076	**0.200**							
bym	-0.323	0.109	-0.744	-0.048	-0.045						
mosse	-0.344	-0.137	-0.706	-0.266	0.071	-0.076					
mtcm	-0.404	-0.119	-0.650	-0.465	0.078	-0.035	0.034				
man	-0.405	-0.068	-0.435	-0.241	**0.171**	0.034	-0.127	-0.031			
defr	-0.170	0.030	-0.462	0.055	**0.217**	-0.009	-0.164	-0.045	-0.090		
campo	**0.151**	**0.151**	-0.089	**0.143**	**0.113**	0.024	0.045	-0.022	0.062	**0.094**	
dbr	-0.155	-0.016	-0.606	0.009	**0.182**	-0.036	-0.141	-0.088	-0.020	**-0.068**	0.034

*Bold values were significant in 16,000 random permutations of the data ($p < 0.10$).

†Sample location abbreviations and map locations listed in Table 2.1 and Figure 2.2.

Analysis of Molecular Variance

Theory and methods

Pairwise estimates of population subdivision are useful summary statistics, especially for assessing overall levels of genetic diversity. However, this approach does not allow for testing hypotheses related to regional genetic differentiation, and its relationship to potential phylogeographic barriers. Michalakis and Excoffier (1996) have adapted the "ϕ-statistics" from the AMOVA approach described in Chapter 3 for use with STR loci. Briefly, AMOVA partitions total genetic variation into different hierarchical levels: within populations, among populations within regions and among regions. In the last partition, regions are exclusive groups of populations that are defined by *a priori* criteria. This property of the AMOVA framework allows a specific geographic structure of the data to be superimposed on a sample population, thereby enabling direct tests of phylogeographic hypotheses about the study population (Excoffier, Smouse, and Quattro 1992; Michalakis and Excoffier 1996).

The AMOVA framework for STR loci is very similar to the AMOVA framework for haploid data presented in Chapter 4. When genotypic data are used in an AMOVA framework an additional level of genetic subdivision can be included: ϕ_{ST} represents genetic division among regions, ϕ_{SC} represents genetic subdivision between groups within regions, and ϕ_{IT} represents genetic differences between individuals within groups. However, differences at the allelic level between individuals, symbolized as ϕ_{IS}, may be included in genotypic AMOVA's as an additional measure of heterozygosity over the entire population being evaluated (Excoffier, Smouse, and Quattro 1992; Michalakis and Excoffier 1996).

Formally, ϕ-statistics are given as the ratio of the variances at different levels of population structure. These ratios are defined as:

5.15) $$\phi_{CT} = \frac{\sigma_a^2}{\sigma_T^2}$$

5.16) $$\phi_{SC} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2 + \sigma_d^2}$$

5.17) $$\phi_{IT} = \frac{\sigma_a^2 + \sigma_b^2 + \sigma_c^2}{\sigma_T^2}$$

5.18) $$\phi_{IS} = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_d^2}$$

where the total molecular variance (σ_T^2) is the sum of the variance due to differences in alleles between populations and between regions (σ_d^2), variance due to differences among alleles within populations (σ_c^2), variance due to differences between genotypes in different populations within regions (σ_b^2), and variance due to genotype differences among populations (σ_a^2) (Schneider *et al.* 1997).

Both the F_{ST} and R_{ST} models may be calculated within the AMOVA framework. For the F_{ST} model, the equations are similar to those presented in Chapter 4. Some modifications are necessary to incorporate the R_{ST} model within the AMOVA framework. For a single locus, the relationship between R_{ST} and ϕ_{ST} may be summarized by:

5.19) $$\phi_{ST} = \frac{S_B - S_W}{S_B}$$

where S_B and S_W are the average squared difference in allele size between pairs of genes from different populations and from the same population, respectively. In order to use the AMOVA framework for multilocus data, a weighted average of the sum of the squared size difference across loci is implemented within the Arlequin program (Slatkin 1995; Michalakis and Excoffier 1996; Schneider *et al.* 1997).

Results and interpretation

Both the F_{ST} and R_{ST} models were applied to the data set within the AMOVA framework. Different potential regional geographic structures were tested including populations on either side of the Sanaga River, the Cross River and the Cameroon Highlands. For the Sanaga River division, samples were divided into two groups: eastern Nigeria and western Cameroon north of the Sanaga River (Figure 2.2, map locations 3 – 12), and Cameroon south and east of the Sanaga River (map locations 13 – 16). For the Cross River division, samples were divided into two groups: west and north of the Cross River (map locations 4 – 9) and south and east of the Cross River (map locations 3, 10 – 16). For the Cameroon Highlands division, samples were divided along the Nigeria-Cameroon border: west of the Cameroon highlands (map locations 3 – 9) and east of the Cameroon Highlands (map locations 10 – 16). All samples from western Nigeria were excluded from these analyses due the possible influence of the Niger River on chimpanzee distribution patterns. Samples sizes for populations in western Nigeria were too small to permit statistical testing of the influence of the Niger River on chimpanzee population structure.

Table 5.3 summarizes AMOVA results for different possible subdivisions of the full multilocus data set. Overall, these results indicate very little subdivision at the regional level for the multilocus data set, and account for 2.24% (R_{ST}) to 2.53% (F_{ST}) of the observed regional variation for the Sanaga River division. Most of the genetic diversity is distributed among chimpanzees at the population and individual levels for each regional partition. Negative ϕ_{CT} values were estimated for both the Cross River (F_{ST}, $\phi_{CT} = -0.012$; R_{ST}, $\phi_{CT} = -0.054$) and Cameroon Highlands (F_{ST}, $\phi_{CT} = -0.031$; R_{ST}, $\phi_{CT} = -0.034$) regional divisions. These regional divisions also produced higher variance components than the Sanaga River division. High variance indicates a loss of statistical power; while negative variance components can indicate an absence of genetic structure (i.e., no genetic differentiation) both within and between populations (Cockerham 1969; Weir 1996a, 1996b). Consequently, the higher variances and negative ϕ_{CT} values associated with the Cross River and Cameroon Highlands divisions make them less likely candidates as phylogeographic boundaries for chimpanzees in Nigeria and Cameroon (Schneider *et al.* 1997). The Sanaga River division, on the other hand, produced positive ϕ_{CT} values in both the F_{ST} and R_{ST} AMOVA's, and in the case of the R_{ST} AMOVA, a much lower variance component (F_{ST}, $\phi_{CT} = 0.026$; R_{ST}, $\phi_{CT} = 0.022$). These findings

Table 5.3. AMOVA for different hypothetical population structures for chimpanzees in Nigeria and Cameroon

Populations were divided into regions based on possible subdivision across different geographic boundaries.

	Conventional F_{ST} genotype frequencies				Standardized R_{ST} sum of squared size difference			
Sanaga River								
Variance Component	Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions σ^2_a	0.040	2.53	>0.120	$\phi_{CT} = 0.026$	18.469	2.24	>0.316	$\phi_{CT} = 0.022$
Among populations/regions σ^2_b	-0.12	0	>1.000	$\phi_{SC} = -0.076$	-41.135	0	>0.797	$\phi_{SC} = -0.051$
Within populations σ^2_c	-0.188	0	<1.00	$\phi_{IT} = -0.162$	366.292	44.34	=0.000	$\phi_{IT} = 0.416$
Within individuals/population σ^2_d	1.819	97.47	>0.100	$\phi_{IS} = -0.109$	482.467	53.42	=0.000	$\phi_{IS} = 0.432$
Cross River								
Variance Component	Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions σ^2_a	-0.018	0	>0.771	$\phi_{CT} = -0.012$	-43.327	0	>0.947	$\phi_{CT} = -0.054$
Among populations/regions σ^2_b	-0.076	0	>0.100	$\phi_{SC} = -0.049$	-12.478	0	>0.586	$\phi_{SC} = -0.015$
Within populations σ^2_c	-0.178	0	<1.000	$\phi_{IT} = -0.178$	395.951	49.63	=0.000	$\phi_{IT} = 0.426$
Within individuals/population σ^2_d	1.802	100	>0.100	$\phi_{IS} = -0.110$	457.629	50.37	=0.000	$\phi_{IS} = 0.464$
Cameroon Highlands								
Variance Component	Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions σ^2_a	-0.047	0	>0.918	$\phi_{CT} = -0.031$	-27.275	0	>0.878	$\phi_{CT} = -0.034$
Among populations/regions σ^2_b	-0.066	0	>1.000	$\phi_{SC} = -0.042$	-21.187	0	>0.675	$\phi_{SC} = -0.025$
Within populations σ^2_c	-0.173	0	<1.000	$\phi_{IT} = -0.188$	399.184	49.38	=0.000	$\phi_{IT} = 0.437$
Within individuals/population σ^2_d	1.802	100	>0.100	$\phi_{IS} = -0.106$	457.629	50.62	=0.000	$\phi_{IS} = 0.466$

[a] Denotes the probability of obtaining an empirical value =,< or > than the null distribution of random values.

suggest that a significant genetic discontinuity at the Sanaga River is more likely than for the other two divisions of the multilocus database.

Null distributions of variances were generated by 16,000 random permutations of the data to test the hypothesis that the observed variance components resulted from sampling error, rather than from real population subdivisioning (Excoffier, Smouse, and Quattro 1992; Michalakis and Excoffier 1996). Representative results are shown in Figure 5.4 from variances calculated for the Sanaga River division of the multilocus database. Open bars represent the null distribution generated by simulation, and closed bars represents the location of the empirical value of each variance component.

Variance among regions is not significantly different from chance, either in the F_{ST} or R_{ST} case. However, in the F_{ST} AMOVA a regional division of the data at the Sanaga River approaches statistical significance, and reveals only a 12% probability that sampling error could account for the genetic differentiation observed at the regional level (Figure 5.4, top panel). There is an absence of genetic structure among populations within regions (Figure 5.4, second panel), implying a high level of gene flow between populations within regions. Among individuals within populations there is a significant excess of genetic diversity (Figure 5.4, third panel). At the gametic phase of individual genotypes (i.e., alleles), there is a statistically significant lack of genetic differentiation (Figure 5.4, bottom panel). The lack of genetic differentiation at the allelic level may be due to problems with allelic dropout (discussed in Chapter 2) that persisted despite multiple attempts to amplify scorable alleles for each individual at each locus via PCR (Gagneux, Boesch, and Woodruff 1997; Goosens, Waits, and Taberlet 1998). In addition, negative variance components may indicate that more alleles are shared

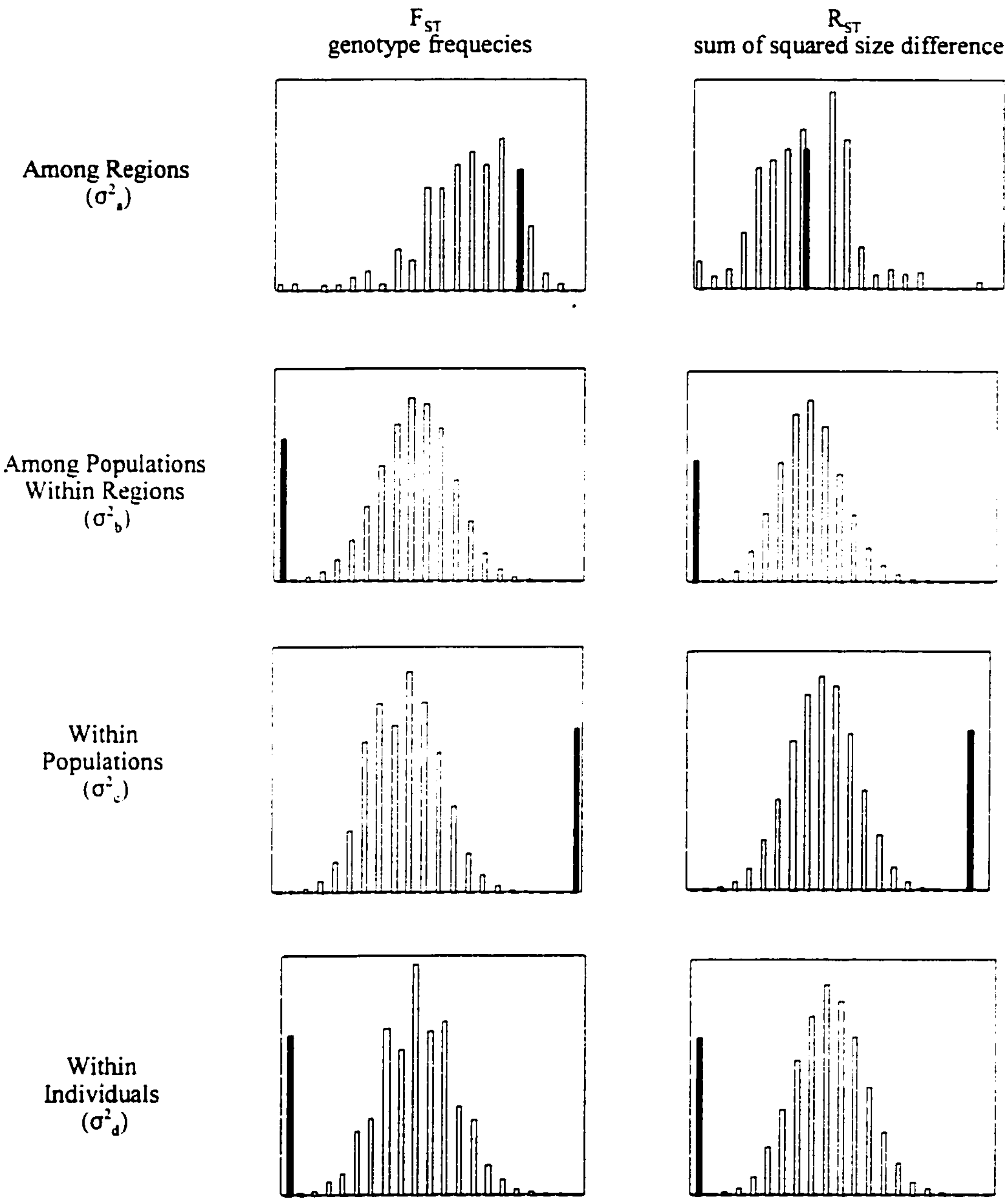


Figure 5.4. Null distributions of F_{ST} and R_{ST} variance components at four hierarchical levels of chimpanzees in Nigeria and Cameroon. Open bars show the distribution of variances generated by 16,000 random permutations of a) populations across regions, b) populations within regions, c) individuals within populations and d) within individuals. Solid markers show the positions of observed empirical values. Two regions were considered: populations north of the Sanaga River and populations south and east of the Sanaga River.

between rather than within populations (Weir 1996a, 1996b). Consequently, the high proportion of shared alleles between individuals may largely explain the high number of negative variance components detected by AMOVA among these chimpanzees.

These findings suggest that chimpanzees in Nigeria and Cameroon are genetically diverse, and are largely panmictic. Moreover, most of the genetic differentiation within the STR multilocus database is restricted to differences between individuals within populations. However, the multilocus AMOVA most strongly supports a regional geographic division between chimpanzees in Nigeria and Cameroon is the Sanaga River, but this division in the multilocus STR database is not statistically significant.

Single locus AMOVA

Bossart and Prowell (1998) suggest that presenting multilocus treatments of STR data can be biologically misleading for two reasons. First, uninformative loci can be introduced into a multilocus analysis. Uninformative loci can obscure signatures of population differentiation or expansion. Secondly, loci that violate neutrality and/or equilibrium assumptions can introduce a large amount of error into any population subdivision analysis. For these reasons, Bossart and Prowell (1998) recommend calculating population subdivision estimates for each locus in a multilocus data set separately. When STR loci are analyzed individually it is possible to distinguish the relative effects of locus-specific evolutionary forces, to determine the relationship of each locus to overall allelic diversity, to evaluate violations of model assumptions and to compare observed patterns of genetic structure.

Therefore, I analyzed each locus separately for each potential phylogeographic division of the STR data, using both the F_{ST} and R_{ST} models. Table 5.4 summarizes single-locus AMOVA's executed for the Sanaga River division, because the Sanaga is the most likely biogeographic boundary to chimpanzee distribution in my study region. Seven of the ten loci reveal that most of the genetic differentiation of the Nigeria-Cameroon chimpanzee population is attributable to differences between individual genotypes. These loci included di-, tri-, tetra- and more complex nucleotide repeat motifs. Three loci (D7, D9 and D11) reveal significant regional subdivision at the Sanaga River when the data were permuted by AMOVA using the R_{ST} model.

The D7, D9 and D11 loci all have tetra-nucleotide repeat motifs. Chakraborty *et al.* (1997) reported that STR's with tetra-nucleotide repeat motifs evolve 1.5 to 2.5 times more slowly than di- or tri- nucleotide repeats. My study examined two di-nucleotide repeat STR's, one tri-nucleotide repeat STR, six tetra-nucleotide repeat motifs and one STR with a very complex repeat motif. If the chimpanzees on either side of the Sanaga River diverged distantly in the past, it is possible that allelic homoplasy at the di- and tri-nucleotide STR loci make these populations appear to have a more recent historical relationship than they actually share. Several investigators have found similar results in a variety of organisms (e.g., Blanquer-Maumont and Crouau-Roy 1995; Garza, Slatkin, and Freimer 1995; Garza and Freimer 1996; Doyle *et al.* 1998; Schug *et al.* 1998; Taylor, Sanny, and Breden 1999). This scenario seems likely due to the differences in ϕ-statistics between the different STR loci types.

Table 5.4. AMOVA for 10 STR loci

Populations were divided into two regions, (1) eastern Nigeria and Cameroon north and (2) Cameroon south of the Sanaga River

		Conventional F_{ST} genotype frequencies				Standardized R_{ST} sum of squared differences			
APOA2									
Variance Component		Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions*	σ^2_a	0.000	0.07	>0.317	$\phi_{CT} = 0.001$	-0.630	0	>0.824	$\phi_{CT} = -0.045$
Among populations/regions	σ^2_b	0.015	3.78	>0.008	$\phi_{SC} = 0.037$	2.148	15.29	>0.029	$\phi_{SC} = 0.146$
Within populations	σ^2_c	-0.008	0	<0.430	$\phi_{IT} = 0.018$	5.275	37.56	=0.000	$\phi_{IT} = 0.484$
Within individuals/populations	σ^2_d	0.392	96.22	>0.670	$\phi_{IS} = -0.021$	7.250	47.15	=0.000	$\phi_{IS} = 0.421$
D4S1652									
Variance Component		Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions	σ^2_a	0.004	0.82	>0.151	$\phi_{CT} = 0.008$	-4.467	0	>0.927	$\phi_{CT} = -0.015$
Among populations/regions	σ^2_b	0.006	1.36	>0.250	$\phi_{SC} = 0.013$	-13.172	0	>0.688	$\phi_{SC} = -0.043$
Within populations	σ^2_c	0.129	28.84	=0.000	$\phi_{IT} = 0.310$	95.790	31.77	<0.003	$\phi_{IT} = 0.259$
Within individuals/populations	σ^2_d	0.309	68.98	=0.000	$\phi_{IS} = 0.295$	223.345	68.23	>0.001	$\phi_{IS} = 0.300$
D7S1809†									
Variance Component		Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions	σ^2_a	0.010	2.32	>0.192	$\phi_{CT} = 0.023$	65.265	38.41	>0.003	$\phi_{CT} = 0.384$
Among populations/regions	σ^2_b	0.041	9.68	>0.012	$\phi_{SC} = 0.010$	6.357	3.74	>0.253	$\phi_{SC} = 0.061$
Within populations	σ^2_c	0.130	30.93	=0.000	$\phi_{IT} = 0.429$	64.599	37.99	=0.000	$\phi_{IT} = 0.801$
Within individuals/populations	σ^2_d	0.239	57.07	=0.000	$\phi_{IS} = 0.351$	33.739	19.86	=0.000	$\phi_{IS} = 0.657$

Table 5.4. AMOVA for 10 STR loci (continued)

	Conventional F_{ST} genotype frequencies				Standardized R_{ST} sum of squared differences			
D9S303†								
Variance Component	Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions σ^2_a	0.005	1.14	>0.243	$\phi_{CT} = -0.011$	26.100	15.32	>0.027	$\phi_{CT} = 0.153$
Among populations/regions σ^2_b	0.005	1.17	>0.245	$\phi_{SC} = 0.012$	-16.614	0	>0.910	$\phi_{SC} = -0.115$
Within populations σ^2_c	0.167	36.55	=0.000	$\phi_{IT} = 0.389$	118.268	59.64	=0.000	$\phi_{IT} = 0.749$
Within individuals/populations σ^2_d	0.277	61.15	=0.000	$\phi_{IS} = 0.374$	42.667	25.04	=0.000	$\phi_{IS} = 0.735$
D11S1984†								
Variance Component	Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions σ^2_a	0.020	4.33	>0.110	$\phi_{CT} = 0.043$	130.679	31.76	>0.040	$\phi_{CT} = 0.317$
Among populations/regions σ^2_b	0.037	7.98	>0.087	$\phi_{SC} = 0.083$	-44.879	0	>0.711	$\phi_{SC} = -0.160$
Within populations σ^2_c	0.289	63.08	=0.000	$\phi_{IT} = 0.754$	224.548	43.65	<0.000	$\phi_{IT} = 0.754$
Within individuals/populations σ^2_d	0.113	24.61	=0.000	$\phi_{IS} = 0.719$	101.162	24.59	>0.015	$\phi_{IS} = 0.689$
HUMPLA2A								
Variance Component	Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions σ^2_a	-0.002	0	>0.580	$\phi_{CT} = -0.007$	-0.641	0	>0.692	$\phi_{CT} = -0.025$
Among populations/regions σ^2_b	0.007	2.17	>0.217	$\phi_{SC} = 0.023$	-0.128	0.76	>0.397	$\phi_{SC} = -0.007$
Within populations σ^2_c	0.177	52.99	=0.000	$\phi_{IT} = 0.552$	21.84	70.60	=0.000	$\phi_{IT} = 0.734$
Within individuals/populations σ^2_d	0.149	44.84	=0.000	$\phi_{IS} = 0.545$	9.093	28.64	=0.000	$\phi_{IS} = 0.718$
D13S317								
Variance Component	Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions σ^2_a	-0.003	0	>0.757	$\phi_{CT} = -0.006$	-0.003	0	>0.754	$\phi_{CT} = -0.005$
Among populations/regions σ^2_b	-0.003	0	>0.570	$\phi_{SC} = -0.006$	-0.003	0	>0.570	$\phi_{SC} = -0.006$
Within populations σ^2_c	0.192	42.57	=0.000	$\phi_{IT} = 0.414$	0.198	42.57	=0.000	$\phi_{IT} = 0.414$
Within individuals/populations σ^2_d	0.264	57.84	=0.000	$\phi_{IS} = 0.421$	0.264	57.84	=0.000	$\phi_{IS} = 0.421$

Table 5.4. AMOVA for 10 STR loci (continued)

	Conventional F_{ST} genotype frequencies				Standardized R_{ST} sum of squared differences			
D16S265								
Variance Component	Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions σ^2_a	-0.009	0	>0.959	$\phi_{CT} = -0.023$	-5.497	0	>0.901	$\phi_{CT} = -0.038$
Among populations/regions σ^2_b	0.027	6.66	=0.000	$\phi_{SC} = 0.065$	15.711	10.85	>0.008	$\phi_{SC} = 0.105$
Within populations σ^2_c	-0.067	0	<0.976	$\phi_{IT} = -0.123$	-36.021	0	<0.908	$\phi_{IT} = -0.178$
Within individuals/populations σ^2_d	0.452	93.34	>0.992	$\phi_{IS} = -0.174$	170.581	89.15	>0.978	$\phi_{IS} = -0.267$
D16S539								
Variance Component	Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions σ^2_a	0.000	0.13	>0.338	$\phi_{CT} = 0.001$	-0.17	0	>0.822	$\phi_{CT} = -0.006$
Among populations/regions σ^2_b	-0.003	0	>0.700	$\phi_{SC} = -0.010$	-0.366	0	>0.676	$\phi_{SC} = -0.013$
Within populations σ^2_c	0.009	2.79	<0.327	$\phi_{IT} = 0.019$	-3.554	0	<0.920	$\phi_{IT} = -0.145$
Within individuals/populations σ^2_d	0.299	97.08	>0.288	$\phi_{IS} = 0.028$	32.235	100	>0.891	$\phi_{IS} = -0.124$
D20S470								
Variance Component	Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions σ^2_a	-0.002	0	>0.564	$\phi_{CT} = -0.004$	-23.370	0	>0.831	$\phi_{CT} = -0.026$
Among populations/regions σ^2_b	0.009	1.98	>0.080	$\phi_{SC} = 0.020$	11.773	1.32	>0.353	$\phi_{SC} = 0.013$
Within populations σ^2_c	0.113	24.54	=0.000	$\phi_{IT} = 0.262$	671.960	72.57	=0.000	$\phi_{IT} = 0.739$
Within individuals/populations σ^2_d	0.340	73.48	=0.000	$\phi_{IS} = 0.249$	233.340	26.11	=0.000	$\phi_{IS} = 0.742$

†Denotes significant differentiation at the regional level.

[a] Denotes the probability of obtaining an empirical value =,< or > than the null distribution of random values.

Migration rates

Gene flow between populations or across regions is notoriously difficult to estimate accurately (Slatkin 1994a). However, Wright (1951) showed that for neutral alleles in an island population structure model that F_{ST} is approximately equal to 1/(1+4Nm), where *N* is the size of the subpopulations and *m* is the average immigration rate. By inverting Wright's formula Nm can be estimated from F_{ST} values by:

5.20)
$$Nm = \frac{1 - F_{ST}}{4F_{ST}}$$

This equation was applied to the regional and population level divisions of the F_{ST} AMOVA's. These calculations suggest that roughly 11 migrants have been exchanged between populations on either side of the Sanaga River per generation. An average of three migrants have been exchanged per generation between populations found only north or south of the Sanaga River.

Migration rates may also be estimated for R_{ST} values. In an island model of finite sample size Slatkin (1995) proposed estimating Nm by:

5.21)
$$M = \frac{1}{4}\left(\frac{1}{F_{ST}} - 1\right)$$

Applying equation 5.21 to the Sanaga River regional and population level R_{ST} AMOVA values, suggests that an average of 11 migrants have been exchanged across the Sanaga

River per generation; and that four migrants have been exchanged between populations found only north or south of the Sanaga River per generation.

These migration rates are high and suggest considerable gene flow across the Sanaga and between populations found only north or south of the river. However, frequent allelic homoplasy in a STR database can greatly decrease F_{ST} and R_{ST} values (Goldstein and Pollock 1997; Bossart and Prowell 1998), and therefore, increase migration rate estimates. It is likely that there is a high level of allelic homplasy in the STR database (discussed on Pp. 202-205). Consequently, these migration rate estimates may be high due to the model assumptions and the quality of the STR data.

Comparisons to other chimpanzees

The apportionment of genetic diversity among sampling locations within Nigeria and Cameroon is complex. The STR database suggests considerable gene flow between populations, and a possible regional division of chimpanzees across the Sanaga River. However, it is unclear how regional chimpanzee populations across Africa differ from one another. Gagneux (1998) and Morin (1992) published STR allele sizes for several chimpanzee populations. Gagneux (1998) genotyped 89 chimpanzees located in Upper Guinea. Morin (1992) provided STR allele sizes of 29 eastern African chimpanzees sampled at Gombe in Tanzania. Three loci used to genotype samples from those studies were also used for this study. Those loci are: APOA2 (repeat unit, AC), HUMPLA2A (repeat unit, AAT) and D16S265 (repeat unit, AC). The same analyses used for the Nigeria-Cameroon STR database were used to examine the apportionment of genetic diversity of chimpanzees across Africa. However, all analyses of chimpanzee populations on a continental level should be interpreted very cautiously, especially for these three loci because they evolve very rapidly (Garza, Slatkin, and Freimer 1995; Rubinsztein *et al.* 1995; Rubinsztein, Leggo, and Amos 1995; Chakraborty *et al.* 1997; Goldstein and Pollock 1997; Schug *et al.* 1998). Consequently, there is a high probability of allelic homoplasy at these three loci.

Population pairwise estimates of genetic differentiation

Pairwise F_{ST} and R_{ST} statistics were used to evaluate genetic differentiation among chimpanzees on a continental level for three STR loci that overlapped with previous studies. Tables 5.5 and 5.6 summarize the pairwise population F_{ST} and R_{ST} statistics for each population. All statistics were calculated with the computer programs Arlequin (Schneider *et al.* 1997) and MICROSAT (Minch *et al.* 1996).

The pairwise comparisons do not reveal any statistically significant pairwise genetic subdivision among chimpanzees limited to Upper Guinea. Pairwise F_{ST} estimates of this population range from –0.029 to 0.021; and R_{ST} values range from –0.157 to 0.034. Among chimpanzees in Nigeria and Cameroon, F_{ST} values range –0.281 to 0.159; and R_{ST} estimates ranged from –0.543 to 0.550. Nine statistically significant pairwise differences occur among sampling locations in Nigeria and Cameroon, but none are related to geographic distance or to obvious population isolation. The pooled pairwise F_{ST} estimate across the continent is 0.012, and is not significantly different from the 16,000 random permutations of the data. The pooled continental R_{ST} estimate is 0.118, and is significantly different from the null distribution of differentiation values ($p < 0.05$).

Table 5.5. Pairwise F_{ST} values of three STR loci for chimpanzee populations across Africa

	TaiA†	TaiB†	Tenkere†	Solo†	ise‡	ggnp‡	ngny‡	akzn‡	crnp‡	bym‡	mosse‡	mtcm‡	man‡	defr‡	campo‡	dbr‡
TaiB	0.012															
Tenkere	0.004	-0.016														
solo	-0.014	-0.013	-0.067													
ise	0.021	-0.029	**0.091***	-0.093												
ggnp	**0.098**	**0.072**	**0.130**	0.003	0.014											
ngny	-0.013	-0.090	0.050	-0.162	0.024	-0.030										
akzn	**0.194**	**0.195**	**0.229**	**0.177**	0.034	0.011	-0.106									
crnp	-0.185	-0.379	-0.115	-0.659	-0.197	-0.106	0.023	-0.466								
bym	0.164	**0.157**	**0.188**	**0.143**	-0.044	**0.152**	0.058	0.049	-0.407							
mosse	**0.176**	**0.180**	**0.199**	**0.166**	0.031	0.032	-0.141	0.023	-0.357	**0.167**						
mtcm	**0.156**	**0.160**	**0.143**	**0.174**	-0.122	-0.048	-0.328	0.037	-0.766	**0.084**	-0.005					
man	**0.045**	**0.038**	**0.127**	0.022	0.021	0.037	0.018	0.001	-0.165	0.075	0.038	-0.109				
defr	**0.195**	**0.223**	**0.242**	**0.209**	0.016	**0.074**	-0.081	-0.017	-0.301	**0.159**	0.006	-0.016	0.046			
campo	-0.028	-0.120	0.031	-0.250	-0.015	0.019	0.035	-0.093	-0.023	0.040	-0.060	-0.281	0.013	-0.041		
dbr	**0.097**	**0.080**	**0.171**	0.030	0.024	0.051	-0.004	-0.004	-0.131	0.052	0.039	-0.114	0.046	-0.041	0.000	
gombe§	**0.250**	**0.235**	**0.215**	**0.196**	**0.149**	**0.120**	-0.074	**0.189**	-0.127	**0.216**	**0.121**	**0.133**	**0.138**	**0.159**	0.018	**0.143**

*Bold values were significant in 16,000 random permutations of the data ($p < 0.05$).

†Populations in Upper Guinea are labeled as TaiA, TaiB, Tenkere and Solo. Data taken from Gagneux (1998).

‡Sample location abbreviations and map locations listed in Table 2.1 and Figure 2.2

§The eastern African population is labeled as Gombe. Data taken from Morin (1992). Allele sizes for this population were given in Gagneux (1998).

Three STR loci were calculated in this matrix: APOA2, D16S265 and HUMPLA2A.

Table 5.6. Pairwise R_{ST} values of three STR loci for chimpanzee populations across Africa

	TaiA†	TaiB†	Tenkere†	solo†	isc‡	ggnp‡	ngny‡	akzn‡	crnp‡	bym‡	mosse‡	mtcm‡	man‡	defr‡	campo‡	dbr‡
TaiB	0.003															
Tenkere	-0.003	0.034														
solo	-0.025	0.033	-0.157													
ise	**0.520***	**0.526**	**0.449**	**0.251**												
ggnp	**0.567**	**0.633**	**0.543**	**0.427**	-0.071											
ngny	**0.319**	**0.465**	**0.377**	0.105	-0.036	0.065										
akzn	**0.557**	**0.621**	**0.528**	**0.412**	0.134	0.160	-0.241									
crnp	**0.521**	**0.683**	**0.592**	**0.419**	-0.460	-0.543	**0.373**	-2.597								
bym	**0.725**	**0.784**	**0.701**	**0.696**	-0.157	0.109	**0.550**	**0.419**	-0.852							
mosse	**0.557**	**0.528**	**0.412**	**0.343**	-0.217	-0.163	-0.210	**0.203**	-0.460	-0.056						
mtcm	**0.562**	**0.480**	**0.297**	**0.315**	-0.260	-0.183	-0.401	0.073	-0.493	-0.258	0.033					
man	**0.631**	**0.668**	**0.603**	**0.497**	-0.052	0.049	0.199	**0.221**	-1.289	0.105	-0.108	-0.131				
defr	**0.606**	**0.611**	**0.483**	**0.454**	-0.190	-0.078	-0.095	**0.196**	-0.419	-0.050	-0.015	-0.106	0.009			
campo	**0.545**	**0.671**	**0.580**	**0.453**	-0.237	-0.098	0.159	-0.443	-0.231	**0.340**	-0.495	-0.927	-0.061	-0.316		
dbr	**0.499**	**0.597**	**0.475**	**0.395**	0.063	**0.193**	-0.086	0.061	-0.499	**0.418**	0.085	-0.178	**0.273**	0.077	-0.042	
gombe§	**0.368**	**0.242**	**0.186**	**0.220**	**0.181**	**0.118**	-0.433	**0.198**	-0.138	**0.186**	**0.301**	**0.321**	**0.197**	**0.215**	-0.450	0.005

*Bold values were significant in 16,000 random permutations of the data ($p < 0.05$).

†Populations in Upper Guinea are labeled as TaiA, TaiB, Tenkere and Solo. Data taken from Gagneux (1998).

‡Sample location abbreviations and map locations listed in Table 2.1 and Figure 2.2

§The eastern African population is labeled as Gombe. Data taken from Morin (1992). Allele sizes for this population were given in Gagneux (1998).

Three STR loci were calculated in this matrix: APOA2, D16S265 and HUMPLA2A.

On a regional level, populations in Upper Guinea, in Nigeria and Cameroon and in eastern Africa are, in most comparisons, significantly differentiated ($p < 0.05$). All sampling locations are significantly differentiated from the eastern African Gombe population in each matrix, except for certain sampling locations from Nigeria and Cameroon, including, Ngel Nyaki Forest Reserve (NGNY), Cross River National Park (CRNP), Campo Forest Reserve and Dja Biosphere Reserve (DBR). In the case of Ngel Nyaki (NNFR), the lack of significant genetic differentiation may be due to local population variances that are larger than the total population variances because of the small sample size from that location. In the Cross River samples (CRNP), two different "central African" mtDNA sequences are present. Perhaps the negative F_{ST} and R_{ST} pairwise values of the Cross River and Gombe populations reflect the presence of "central African genotypes" within the Cross River samples. The southern Cameroon populations (Campo and Dja) may, however, share a more recent historical relationship with more easterly chimpanzee populations than with those further to the west.

Analysis of Molecular Variance

The AMOVA framework was also applied to the pan-African STR database using the computer program Arlequin (Schneider *et al.* 1997). ϕ-statistics were calculated for the three loci combined and for each locus separately. The significance of each ϕ-statistic was evaluated against 16,000 random permutations of the data. Results from the F_{ST} and R_{ST} AMOVA's are summarized in Table 5.7. Comparisons of null and observed variances at each hierarchical level for the multilocus AMOVA's are shown in Figure 5.5.

Table 5.7. AMOVA for different hypothetical population structures for chimpanzees across Africa

Populations were divided into three regions, Upper Guinea, Nigeria and Cameroon, and eastern Africa

		Conventional F_{ST} genotype frequencies				Standardized R_{ST} sum of squared size difference			
ALL LOCI									
Variance Component		Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions	σ^2_a	0.158	16.45	>0.000	$\phi_{CT} = 0.165$	88.360	44.46	>0.000	$\phi_{CT} = 0.445$
Among populations/regions	σ^2_b	-0.016	0	>0.894	$\phi_{SC} = -0.020$	-5.95	0	>0.964	$\phi_{SC} = -0.053$
Within populations	σ^2_c	-0.078	0	<0.452	$\phi_{IT} = 0.066$	-26.310	0	<0.019	$\phi_{IT} = 0.282$
Within individuals/population	σ^2_d	0.896	83.55	=1.000	$\phi_{IS} = -0.095$	142.640	55.54	>0.996	$\phi_{IS} = -0.226$
APOA2									
Variance Component		Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions	σ^2_a	0.090	18.83	=0.000	$\phi_{CT} = 0.188$	45.011	68.14	>0.001	$\phi_{CT} = 0.681$
Among populations/regions	σ^2_b	0.012	2.45	>0.003	$\phi_{SC} = 0.030$	1.222	1.85	>0.013	$\phi_{SC} = 0.058$
Within populations	σ^2_c	-0.014	0	=0.000	$\phi_{IT} = 0.184$	-0.316	0	=0.000	$\phi_{IT} = 0.695$
Within individuals/population	σ^2_d	0.391	78.72	>0.797	$\phi_{IS} = -0.036$	20.144	30.01	>0.578	$\phi_{IS} = -0.016$
HUMPLA2A									
Variance Component		Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions	σ^2_a	0.110	24.29	>0.001	$\phi_{CT} = 0.243$	34.289	58.71	>0.002	$\phi_{CT} = 0.581$
Among populations/regions	σ^2_b	0.011	2.40	>0.145	$\phi_{SC} = 0.032$	1.476	2.53	>0.301	$\phi_{SC} = 0.061$
Within populations	σ^2_c	0.083	18.48	=0.000	$\phi_{IT} = 0.452$	8.252	14.13	=0.000	$\phi_{IT} = 0.754$
Within individuals/population	σ^2_d	0.248	54.83	=0.000	$\phi_{IS} = 0.252$	14.386	24.63	=0.000	$\phi_{IS} = 0.365$
D16S265									
Variance Component		Variance	% Total	p^a	ϕ-statistics	Variance	% Total	p^a	ϕ-statistics
Among regions	σ^2_a	0.059	12.16	>0.000	$\phi_{CT} = 0.122$	52.386	28.93	>0.008	$\phi_{CT} = 0.289$
Among populations/regions	σ^2_b	0.014	2.83	>0.001	$\phi_{SC} = 0.032$	5.070	2.80	>0.021	$\phi_{SC} = 0.039$
Within populations	σ^2_c	-0.031	0	>0.030	$\phi_{IT} = 0.086$	-22.233	0	<0.111	$\phi_{IT} = 0.195$
Within individuals/population	σ^2_d	0.445	85.01	>0.994	$\phi_{IS} = -0.075$	145.867	68.27	>0.961	$\phi_{IS} = -0.180$

[a] Denotes the probability of obtaining an empirical value =,< or > than the null distribution of random values.

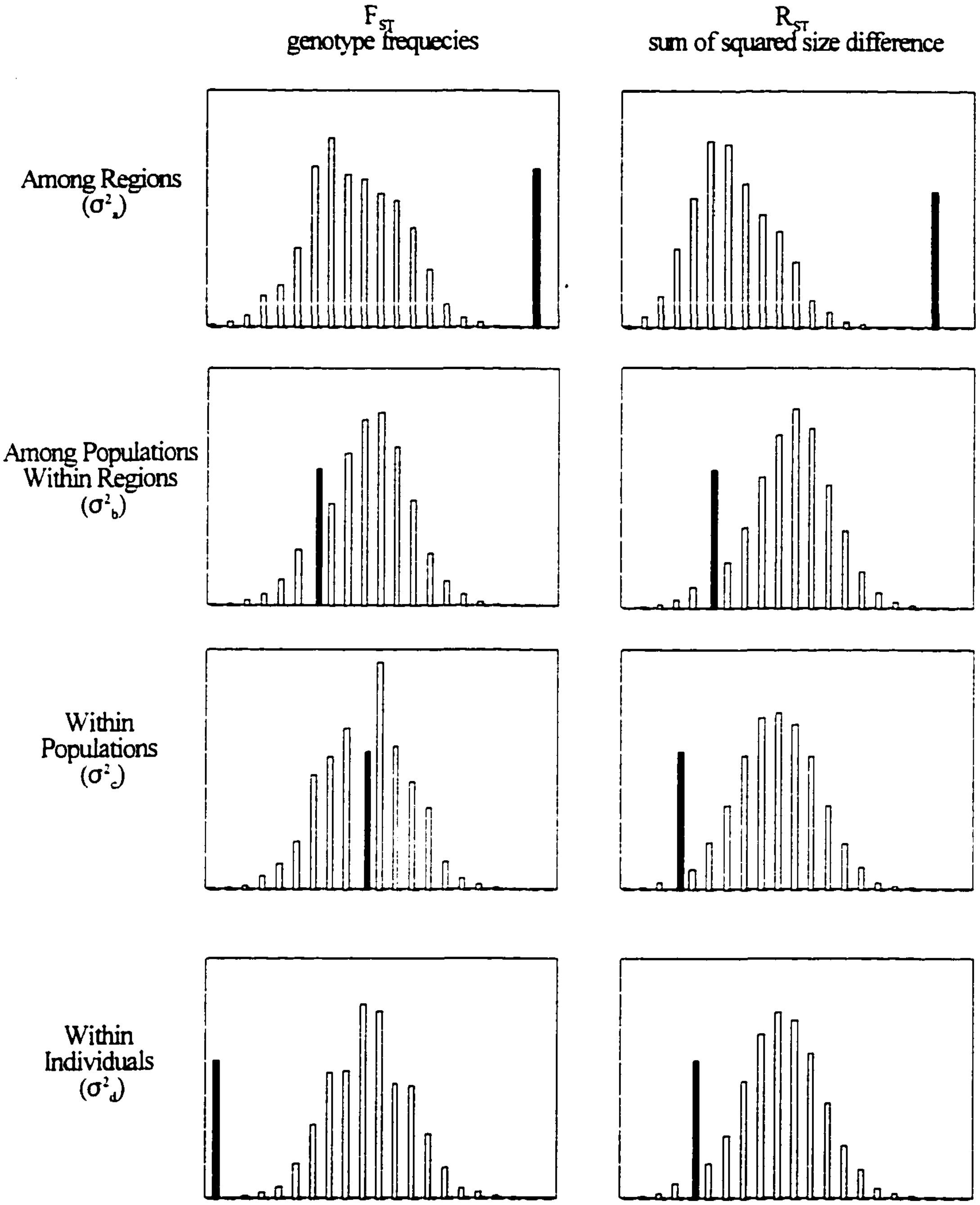


Figure 5.5. Null distributions of F_{ST} and R_{ST} variance components at four hierarchical levels of chimpanzee populations across Africa. Open bars show the distribution of variances generated by 16,000 random permutations of a) populations across regions, b) populations within regions, c) individuals within populations and d) within individuals. Solid markers show the positions of observed empirical values. Three regions were considered: Upper Guinea, Nigeria and Cameroon, and eastern Africa.

Both the multilocus and single-locus AMOVA's indicate a significant excess of genetic differences between regional chimpanzee populations. Regional genetic differences account for 16.45% (F_{ST}, $\phi_{CT} = 0.165$) to 44.46% (R_{ST}, $\phi_{CT} = 0.445$) of the total sample variance. Populations within regions do not contribute to the observed genetic differences at the STR loci in either multilocus AMOVA. However, the multilocus R_{ST} does suggest a statistically significant lack of genetic differentiation between populations within regions. The R_{ST} and F_{ST} multilocus AMOVA's reveal a statistically significant lack of genetic differentiation at the allelic level between individual chimpanzees. In fact, 93.36% (F_{ST}) and 71.77% (R_{ST}) of the genetic diversity among chimpanzees across Africa is attributable to differences between individuals at the allelic level. These results suggest that although there is significant subdivision at the regional level, the bulk of genetic diversity among chimpanzees is attributable to differences between individuals, and that within regions, chimpanzee populations are largely panmictic.

The results of the single-locus AMOVA's are very similar to the multilocus AMOVA's. Levels of genetic differentiation at the regional level are significantly different for all three loci ($p<0.05$). Depending on the locus and mutational model used, regional variation accounts for 12% to 68% of the total observed variance. Each single-locus AMOVA reveals a statistically significant lack of genetic differentiation between populations within regions. Conversely, a significant excess of genetic differences between individuals within populations is present in each single-locus AMOVA. Except for the APOA2 and HUMPLA2A R_{ST} AMOVA's, these analyses suggest that the bulk of the genetic differentiation on a continental level is attributable to differences between individuals, despite the fact that significant regional subdivision exists within the database.

Migration rates

Migration rates between regions were calculated using equations 5.20 and 5.21. Unsurprisingly, these rates were low. The F_{ST} multilocus regional ϕ_{CT} statistic suggests that roughly one migrant per generation has been exchanged between regional chimpanzee populations. The R_{ST} multilocus regional ϕ_{CT} statistic reveals an even lower estimated migration rate of less than one individual migrating between regions per generation ($Nm = 0.32$). At the population level, the multilocus F_{ST} and R_{ST} AMOVA's reveals a much higher average. Between, 4 (R_{ST}) and 12 (F_{ST}) migrants have been exchanged between regions and between populations within regions per generation. The lower migration rates calculated by R_{ST} are attributable to the fact that the R_{ST} ϕ-statistics are larger than the F_{ST} ϕ-statistics. These findings suggest substantial population genetic differentiation at the regional level, but relatively high levels of gene flow between chimpanzee populations within regions.

Comparisons to the mtDNA HV-I sequence database

There are a number of striking similarities and contrasts between the patterns of variation in mtDNA HV-I sequences and STR genotypes both within Nigeria and Cameroon and among chimpanzee populations on a continental level. Both databases document a high level of genetic diversity among chimpanzees in Nigeria and Cameroon, as well as across the African continent. The mtDNA database suggests that chimpanzee populations are divided at the Sanaga River into two ancient clades: a western African lineage and a central African lineage. Within those two ancient clades, both the parsimony tree and the jackknife neighbor-joining tree delineate the differences between chimpanzees in Upper Guinea from those in Nigeria and western Cameroon. Yet, the same analyses do not reliably verify the differences between western equatorial and eastern African chimpanzees, suggesting a much more recent relationship between these populations.

Within the STR database, significant differences are present between chimpanzee populations in Upper Guinea, in Nigeria and Cameroon and in eastern Africa at Gombe. However, the continental STR database does not provide the same level of geographic coverage that the mtDNA database does, making the STR database suspect in providing unbiased estimates of population differentiation. Moreover, three different investigators, using different laboratory and analytical protocols determined the allele sizes of the loci used to compare chimpanzee populations on a continental level. Consequently, differences in allele scoring criteria between studies may have been so different that shared or unique alleles among chimpanzees on a continental level may have been overlooked in the analyses presented in this chapter.

The mtDNA database provides strong statistical support for the proposition that the Sanaga River divides chimpanzee populations into two genetically distinct groups. In contrast, the STR database implies a relatively panmictic population in Cameroon with only a weak differentiation at the Sanaga River. In addition, the ϕ-statistics of the AMOVA's are much lower and the migration rate estimates are much higher for the STR database than in the mtDNA database. The STR database suggests that an average of 11 migrants have been exchanged between populations across the Sanaga River per generation; while the mtDNA database suggested only one migrant per generation has been exchanged on average between populations on either side of the Sanaga River. Finally, both databases suggest that there have been high rates of gene flow and migration among populations restricted to either the north or south of the Sanaga River.

Limitations of the STR database

There are several reasons why these discrepancies may exist between the two databases. First, the stepwise mutation process operating on STR loci is not well understood. Substantial evidence suggests that the number of repeat units attainable by STR loci is restricted, such that a given locus can only contain a certain maximum number of repeats (Goldstein and Pollock 1997). Confounding this problem is that the mutational process itself seems to degrade over time so that if two divergent lineages share alleles at a given locus, that locus may be under different evolutionary constraints for each lineage, making the alleles incomparable (Garza and Freimer 1996; Goldstein and Pollock 1997). Adding to the problem of mutation process decay is that by using heterologous primers, even between closely related species, pure stretches of repeats will often be interrupted with imperfections (Garza, Slatkin, and Freimer 1995; Crouau-Roy *et al.* 1996; Garza and Freimer 1996).

If there are constraints on maximum allele size, on the longevity of the mutational properties and on the purity of repeat series between lineages at STR loci, then the accuracy of distance measures that assume time linearity, such as R_{ST}, may be strongly affected (Slatkin 1995; Goldstein and Pollock 1997; Taylor, Sanny, and Breden 1999).

Consequently, it is possible that there is homoplasy in the STR database; and that the mutational process may not follow a step-wise mutation model. It is interesting that three of the more conserved tetra-nucleotide repeat loci used in this study suggested a significant division of chimpanzees at the Sanaga River. If the phylogeographic division at the Sanaga River is a real and ancient phenomenon, it is possible that the more rapidly-evolving di- and tri- nucleotide repeat STR loci have obscured signatures of regional subdivision due to high levels of allelic homoplasy.

There is some evidence to suggest that there are constraints on maximum allele size in the chimpanzee STR database. General information about human size variation at these STR loci used during this can be accessed at http:\\www.chlc.org. At the APOA2, D16S265, HUMPLA2A and D13S317 loci, several chimpanzees possess alleles that occur at the upper range of the allele sizes reported in humans, and in some cases, exceed the reported sizes. Unfortunately, the number of human genotypes available for analysis was too small to permit statistical testing of the hypothesis that there are constraints on the STR allele sizes in chimpanzees at the loci used for this study. Moreover, without directly sequencing human and chimpanzee putative STR allele homologues, there is no way to be certain that the differences detected in allele size are directly attributable to repeat unit size variation.

Several of the DNA extracts failed to produce scoreable allele sizes at six of the ten loci, despite repeated attempts to amplify them via PCR. It is likely that part of the reason that the STR database does not provide significant support of a regional division of chimpanzees at the Sanaga River is that the data are fragmented. In some cases, the sampling variance of subpopulations exceeded that of the total population. As a result, the negative F_{ST} and R_{ST} values may reflect problems with sample size and with analytical strategy, and not high levels of gene flow or recent historical association (Weir and Cockerham 1984; Nei 1987; Weir 1996a, 1996b). Consequently, the statistics presented in this chapter for these populations may be an inaccurate estimate of the actual amount of differentiation between populations in Nigeria and Cameroon. In order to explore that idea, each analysis presented in this chapter was modified to constrain all populations to an identical size. Although there were some numerical differences after imposing this constraint, the overall results and the significance probabilities of each analysis did not change.

The STR loci chosen for this study may have been inappropriate. This is the first study of the genetics of chimpanzee populations in Nigeria and Cameroon. Because so little is understood about the history and timing of events in this geographical region, it was difficult to predict beforehand what types of loci would have been appropriate to test different phylogeographic hypotheses. Consequently, a wide range of STR loci were chosen for study. Presumably, these loci are subject to different mutational constraints. Consequently, these loci may characterize the relationships between populations in Nigeria and Cameroon on different time scales.

The tetra-nucleotide repeat STR's show the most promise for more clearly delineating the relationships between these populations, but only half of the STR loci used during this study fall into that category. This small number of tetra-nucleotide repeat motif loci renders any final interpretations of the data tenuous at this time. Goldstein *et al.* (1999) suggested 19 different STR loci are required to sufficiently evaluate the demographic history of populations of island foxes (*Urocyon littoralis*).

Following their recommendation, the addition of several more tetra-nucleotide repeat STR loci may more clearly resolve the demographic history of chimpanzees in Nigeria and Cameroon.

Finally, it is possible that analyses of the mtDNA database are misleading. There is some evidence to suggest that at the HV-I locus there may be a high degree of homoplasy (i.e., the variation in transition:tranversion ratios between versus within lineages). However, when compared to the STR database analyses, the inferred demographic histories are largely the same. This observation suggests that the mtDNA database analyses are not misleading. The addition of more samples and more appropriate STR loci may more clearly resolve the similarities and contrasts between the nuclear and mitochondrial genomes.

CHAPTER 6

Discussion

Introduction

The widely accepted view of chimpanzee phylogeography is that this species is divided into three lineages that are geographically separated by the Niger River and by the Ubangi River (Hill 1967, 1969; Goodall 1986; Teleki 1989). Yet, this view of recent chimpanzee evolution does not fit well with the putative history of African forests during the Pleistocene or with the distribution patterns of other forest primates. It is perhaps not surprising, therefore, that the apportionment of genetic diversity found among wild chimpanzees in this study does not coincide well with the widely accepted taxonomy of this species.

Three questions were posed in Chapter 1 in an attempt try to understand these differences, using genetic data obtained from chimpanzees both from Nigeria and Cameroon and from other regions of Africa. These questions were: 1) What is the relationship between chimpanzees in Nigeria and Cameroon and other populations?; 2) What geographic barriers, if any, separate chimpanzee populations in Nigeria and Cameroon from other populations?; and 3) How effectively, and for how long, have dispersal barriers limited chimpanzee distribution in Nigeria and Cameroon?

These questions grew from preliminary work on the topic which suggested that an unrecognized lineage of chimpanzee may occupy Nigeria and adjacent parts of Cameroon

(Gonder *et al.* 1997), and from hypotheses proposed by other researchers (Morin *et al.* 1994) concerning the phylogenetic relationship of chimpanzees in Upper Guinea to other populations. In addition, our preliminary analysis of genetic diversity at the HV-I locus revealed that a significant genetic discontinuity between western and western equatorial African chimpanzees occurs somewhere in central Cameroon. These analyses also suggested that a second phylogeographic boundary may occur between chimpanzees in western Africa at the Dahomey Gap, but not at the lower Niger River in Nigeria. This final chapter addresses these questions by synthesizing the findings of various analyses for the mitochondrial and nuclear DNA databases presented separately in Chapters 3, 4 and 5. This chapter also explains these data within the context of African forest history during the Upper Pleistocene; and concludes with options for further research and possible conservation priorities for chimpanzees in Nigeria and Cameroon.

Chimpanzee phylogenetics and biogeography

At a minimum, the geographical patterning of chimpanzee genetic diversity found in this study suggests that the widely accepted subspecies distributions of this species must be revised. My results suggest two different hypotheses about recent chimpanzee evolution. Figures 6.1 and 6.2 illustrate the geographic distribution patterns of *Pan troglodytes* that correspond to these hypotheses. Figure 6.1 shows a phylogeographic division of chimpanzees in western and western equatorial Africa in the vicinity of the Sanaga River in Cameroon. Figure 6.2a extrapolates that phylogeographic division to include distributions of chimpanzees from other areas of Africa. In this distribution map, two chimpanzee lineages are present in African forests, delimited in their distribution by

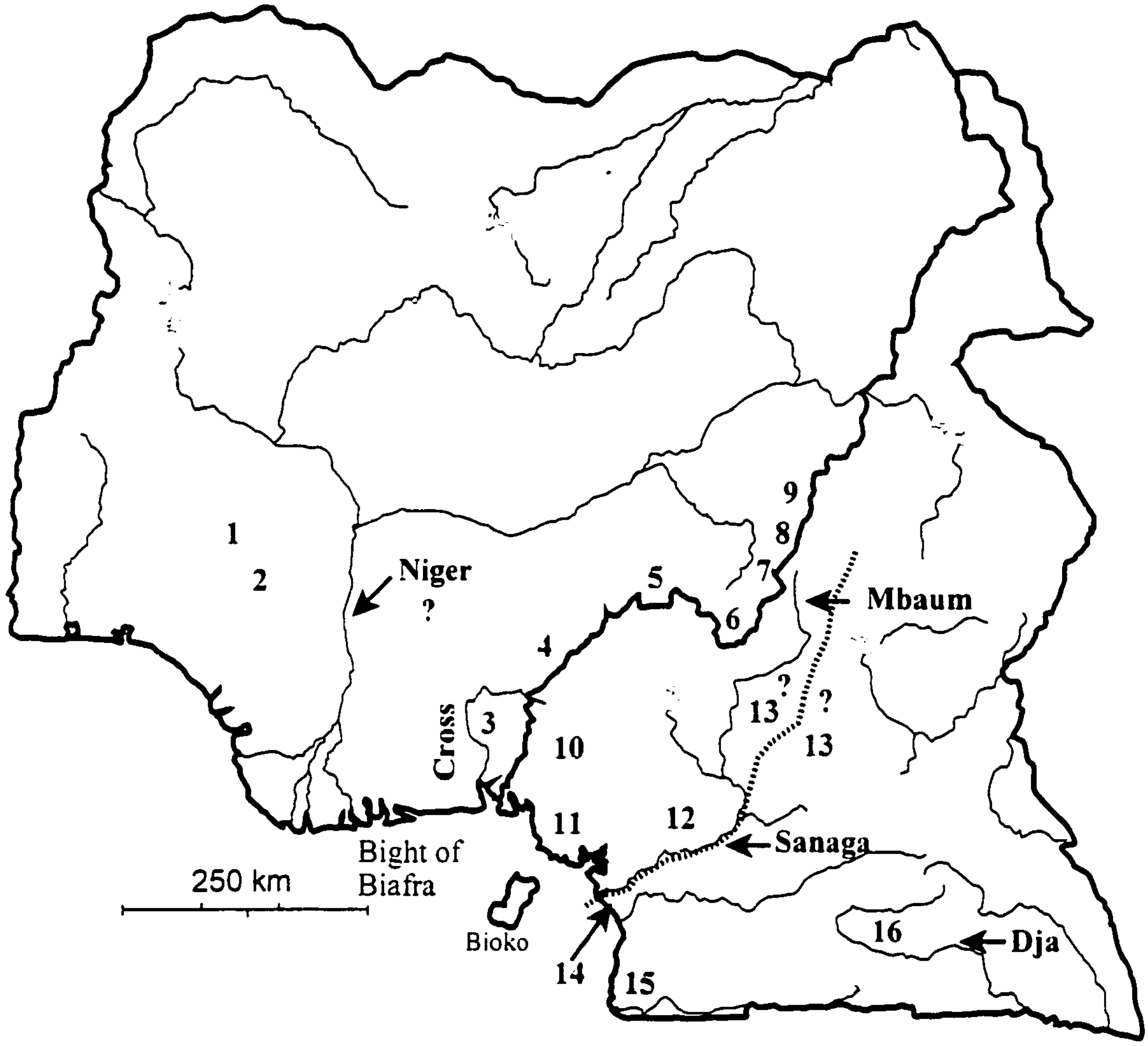


Figure 6.1. Phylogenetic division inferred between chimpanzees in Nigeria and Cameroon. On this map, chimpanzee populations across the study area are divided into two major lineages. The dashed line represents a major phylogeographic division of chimpanzees at the Sanaga River in Cameroon, that appears to have influenced chimpanzee distribution patterns for the last few hundred thousand years. Samples collected from Nigeria and western Cameroon form part of a western African chimpanzee lineage Samples from southern Cameroon represent part of a central African chimpanzee lineage. Sampling location 13 appears on this map twice because evidence of HV-I mtDNA sequences from both lineages are present in the samples collected from that sampling location.

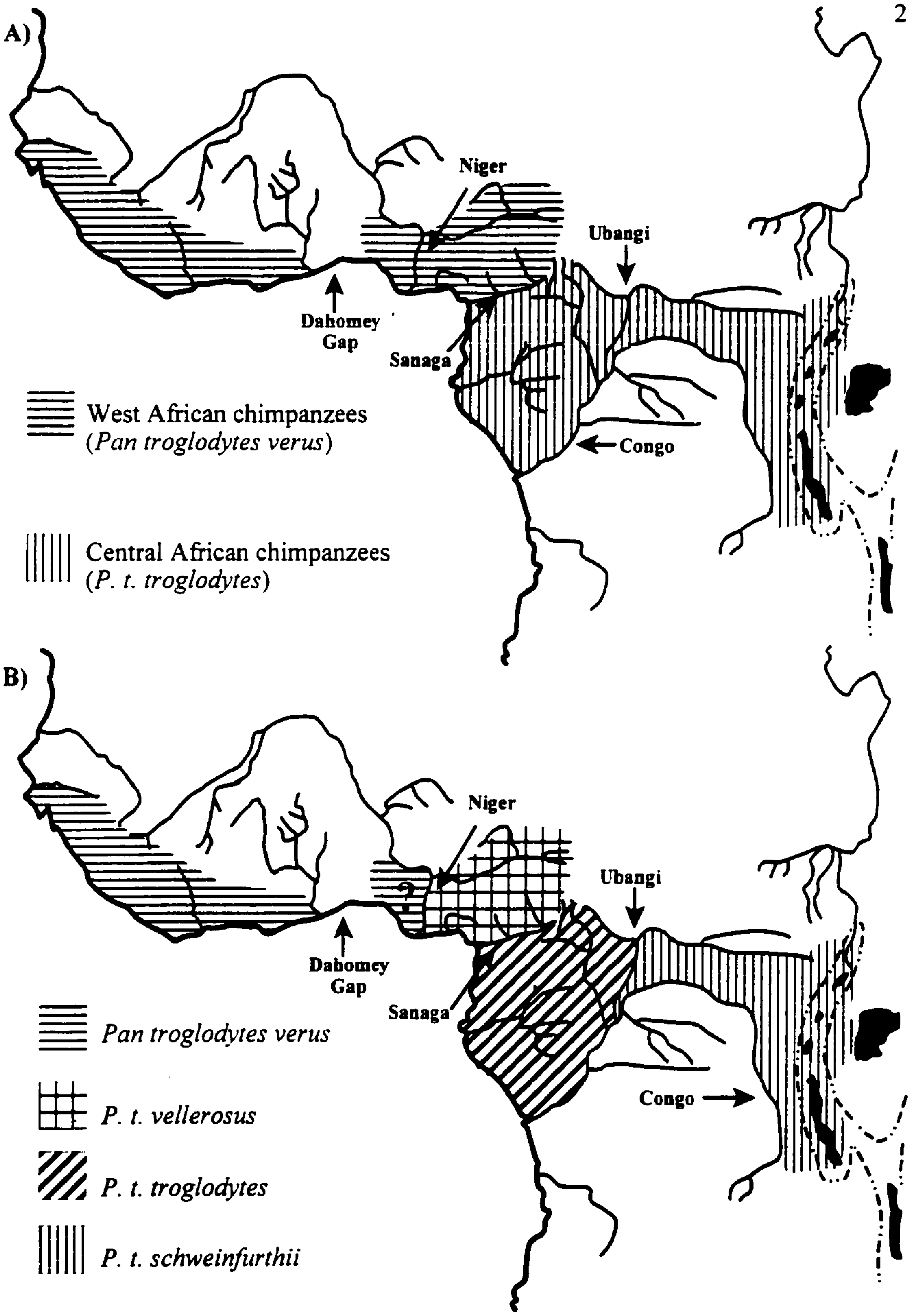


Figure 6.2. Two chimpanzee subspecies arrangements inferred from the pan-African genetic database.

the Sanaga River. Chimpanzees in Upper Guinea, Nigeria and western Cameroon belong to one lineage. The other lineage is composed of chimpanzees in western equatorial and eastern Africa, beginning in southern Cameroon and extending to the eastern perimeter of this species' range. In Figure 6.2b, the western lineage is subdivided further into two separate groups. The lineage represented by chimpanzees in Upper Guinea and western Nigeria is divided into two populations that are geographically separated by the Dahomey Gap. The Niger River separates chimpanzees in western Nigeria from those in eastern Nigeria and western Cameroon, while the Ubangi River separates chimpanzees in western equatorial and eastern Africa.

Each of the phylogenetic reconstructions of the mtDNA HV-I sequence database presented in Chapter 3 suggested that two deeply divergent chimpanzee lineages are present in African forests (Figures 6.1 and 6.2a). Geographically, populations within these clades revealed a deep west-east division. The western African clade was composed mostly of chimpanzees from Upper Guinea, Nigeria and western Cameroon, while the central African clade was composed mostly of chimpanzees from western equatorial and eastern Africa.

The phylogenetic reconstructions suggested also that the distributions of these lineages converge in central Cameroon, with a fairly sharp division of western and central African mtDNA sequences on either side of the Sanaga River. In fact, this genetic discontinuity was so sharply defined that all sequences from Mosse (Figure 5.1, map location 12) and from the Douala-Edea Forest Reserve (map location, 14) belonged to the western and central African clades, respectively. Other analyses of the mtDNA data also supported these findings. For example, the AMOVA and spatial autocorrelation analyses pointed to a major genetic division of chimpanzees in the vicinity of the Sanaga River. The multi-modality of the mismatch distribution for the Nigeria-Cameroon data set suggested that two ancient subdivided lineages are represented in the mtDNA HV-I sequences. When the sequences were divided into two groups restricted to either to the north or south of the Sanaga River, the mismatch distributions suggested a more ancient lineage occupies southern Cameroon; and that a more youthful population occupies eastern Nigeria and western Cameroon.

The migration rate across the Sanaga River has probably been very low. On average, the mtDNA database suggested that only a single migrant has been exchanged each generation between populations on either side of the Sanaga. In contrast, about 26 migrants have been exchanged each generation between chimpanzee populations in forests in the Democratic Republic of Congo and those further to the east (Goldberg 1996). Migrants have been exchanged more frequently each generation between populations restricted to either the north or south of the Sanaga River ($Nm = 2$). Migration rates between populations restricted to one side of the Sanaga were similar to, but lower than, migration rates calculated for other regional chimpanzee populations. Roughly six migrants have been exchanged between chimpanzee populations in Upper Guinea (Gagneux 1998); while roughly three migrants have been exchanged per generation between chimpanzee populations in eastern Africa (Goldberg 1996).

However, there is evidence of mixed populations and of long-range gene flow in Nigeria and Cameroon. mtDNA sequences that are characteristic of both the western and central African clades were detected east of the Mbaum River in eastern Cameroon at Manb'ra (map location, 13). The Mbaum River is the main tributary of the Sanaga River.

The mtDNA database also suggested some long-range gene flow across the Sanaga River. Two "central African" mtDNA sequences were detected in Cross River National Park in southeastern Nigeria (CRNP 6 and CRNP 12; map locations 3 and 4). One "western African" mtDNA sequence was detected in the Dja Biosphere Reserve in southern Cameroon (DBR7; map location, 16). These findings suggest that although most of the genetic differentiation among chimpanzees at the HV-I locus is attributable to differences between populations that occur either north or south of the Sanaga River, the Sanaga may not completely deter gene flow between chimpanzee lineages or that the observed haplotypes must be very ancient.

The relative divergence dates inferred from the gene trees imply that the differences between the western and central African chimpanzee lineages are ancient. The western and central African clades may have originated as early as 900kya, and were well differentiated from one another by at least 200kya. The population expansion dates calculated from the mismatch distributions of the Nigeria-Cameroon chimpanzee mtDNA sequences imply that the Sanaga River, or some feature in its vicinity, has had a long term influence on the apportionment of chimpanzee genetic diversity. Although these estimates ranged widely, the Sanaga River probably has influenced chimpanzee distribution patterns for the last 100 to 200ky, and perhaps even longer. There is a large difference in the relative ages of each lineage calculated by both methods. Chimpanzees in western equatorial Africa carry more diverse, and probably, more ancient mtDNA HV-I sequences than any other population. HV-I sequences from Nigeria and western Cameroon are less diverse, and might signal a more recent expansion of chimpanzees into western Africa.

Analyses of the STR database were ambiguous. Both the mtDNA and STR databases suggested a recent historical relationship and high levels of gene flow between populations on either side of the Sanaga River. In contrast to the mtDNA database, analyses of the STR database suggested that the migration rate across the Sanaga River has been much higher. Migration rate estimates calculated from the STR database imply that approximately 11 migrants have been exchanged per generation between populations on either side of the Sanaga River. Yet despite this high estimate, several divisions and permutations of the STR database by AMOVA pointed to the Sanaga River as the most likely regional divider of chimpanzee populations in Nigeria and Cameroon. However, this regional division of the multilocus STR database was not statistically significant, and accounted for only two percent of the observed genetic differentiation. Moreover, in each permutation of the STR data, the majority of the genetic differences detected were attributable to differences between individuals.

Given the different mutational constraints and processes that may affect different STR loci, each locus was analyzed individually. When each STR locus was analyzed separately by AMOVA, three of the ten loci (D7, D9 and D11) revealed a statistically significant genetic division of chimpanzee populations at the Sanaga River. The D7, D9 and D11 loci are composed of tetra-nucleotide repeats, which evolve more slowly than STR's with di- or tri- nucleotide repeat motifs (Chakraborty *et al.* 1997; Schug *et al.* 1998). If the mtDNA database correctly suggests an ancient division of chimpanzees at the Sanaga River, the more rapidly evolving STR loci may have lost meaningful phylogenetic signal due to allelic homoplasy or due to degradations in the mutational process (Blanquer-Maumont and Crouau-Roy 1995; Estoup *et al.* 1995; Goldstein and Pollock 1997; Schug *et al.* 1998; Taylor, Sanny, and Breden 1999).

Given these potential problems it is surprising that a trend towards the division of the STR database at the Sanaga River was detected at all. The STR markers chosen for this study were picked at random within a few loosely defined criteria (i.e., quality of PCR product, variability, allele scoring consistency and the availability of genotypes from other regional chimpanzee populations). It seems likely that with more genotypes and more loci, especially those with more slowly evolving tetra-nucleotide repeat motifs, that STR genotyping may more clearly resolve the apportionment of chimpanzee genetic diversity in Nigeria and Cameroon.

The relationships between regional chimpanzee populations within the major western and central division of chimpanzees are more ambiguous (Figure 6.2b). mtDNA sequences of chimpanzees in western equatorial and eastern Africa were not well differentiated from one another. None of the phylogenetic reconstructions presented for the mtDNA sequence data suggested that chimpanzees in western equatorial and eastern Africa are two separate lineages. In addition, two "eastern African" chimpanzee HV-I sequences were found in the Cameroon sample populations (BYM3 and Campo9). Although there were no fixed nucleotide substitutions that distinguish chimpanzees in eastern Africa from those in western equatorial Africa, BYM3 and Campo 9 clustered with HV-I sequences of chimpanzees in eastern Africa in the phylogenetic reconstructions. The linearized F_{ST} matrix, which calculates relative dates of coalescence between pairs of populations, implied a more recent relationship between chimpanzees in western equatorial and eastern Africa than between populations in western Africa.

In contrast, the STR database suggested a significant division of western equatorial and eastern African chimpanzees. However, the STR loci for which genotypes of chimpanzees from these populations overlapped [APOA2, HUMPLA2A and D16S265] are known to mutate rapidly (Garza, Slatkin, and Freimer 1995; Chakraborty *et al.* 1997; Schug *et al.* 1998). Allelic homoplasy at these loci probably has made it impossible to determine precisely the relationship between those populations.

The parsimony and jackknife neighbor-joining phylogenetic reconstructions reliably separated chimpanzees in Upper Guinea and western Nigeria from those in eastern Nigeria and western Cameroon into two paraphyletic clades. The maximum-likelihood analysis and the neighbor-joining tree constructed from a distance matrix consisting of only transversion substitutions, however, did not separate these two putative lineages. In contrast, the pairwise linearized F_{ST} matrix suggested that chimpanzees in Upper Guinea (including samples from western Nigeria) and those in Nigeria and western Cameroon share a more distant relationship than populations in western equatorial and eastern Africa. Moreover, STR genotypes of these two populations overlapped at three loci. Analyses of these loci also suggested a significant division of chimpanzees in Upper Guinea from those in eastern Nigeria and Cameroon. However, the same problems of the STR database apply to these comparisons as for those among chimpanzees in western equatorial and eastern Africa.

Unfortunately, the sample size of chimpanzees from western Nigeria was too small to permit statistical testing of the Niger River as a possible biogeographic boundary. Both the parsimony and jackknife neighbor-joining phylogenetic reconstructions of the mtDNA sequences suggested that chimpanzees from western

Nigeria may form part of a clade with chimpanzees limited to Upper Guinea. Alternatively, the maximum-likelihood reconstruction suggested that they may be basal members of the western African clade.

Samples from western Nigeria were collected at Ise Forest Reserve (map location, 1). Ise is a small forest fragment that is bordered on all sides by the Ogbesse River and by cocoa plantations (pers. obs.). It is possible, but very unlikely, that these chimpanzees may be divergent from other chimpanzee populations due to human influence. However, cocoa farming was only introduced to western Africa during the 1800s (Oates 1999). In addition, there is little evidence of widespread agriculture in Africa before 3kya. Therefore, the *maximum* amount of time that these chimpanzees could have been influenced by human activities is on the order of several generations. Consequently, chimpanzees in western Nigeria probably are part of a lineage that also includes chimpanzees limited to Upper Guinea or they may be different from other chimpanzees in western Africa.

Both the mtDNA and STR databases suggest that either the Niger River or the Dahomey Gap have, to some degree, influenced the apportionment of genetic diversity among chimpanzees in western Africa. Although the evidence more strongly supports a division of western African chimpanzees at the Niger River, the fragmentary nature of the data obtained from chimpanzees in western Nigeria makes it impractical to rank-order the importance of either putative biogeographic barrier. Whichever barrier ultimately proves to be of greater importance in recent chimpanzee evolution, these data suggest that something is acting to produce more population substructure within the western chimpanzee lineage than within the central chimpanzee lineage.

Chimpanzee distribution and nomenclature

The data presented in this dissertation most strongly support the distribution and nomenclature model depicted in Figure 6.2a. The overall picture of the data suggests that two deeply divergent lineages of chimpanzees are present in African forests. The ranges of these lineages converge in central Cameroon; and it is likely that the Sanaga River has played an important, but perhaps not an exclusive, role in limiting gene flow between lineages. This genetic differentiation has probably existed between chimpanzees for the past 100 to 200ky, and probably for much longer. However, these lineages have not been completely isolated from each other.

The relationships between chimpanzees within the western and central lineages are complex. Across western equatorial and eastern Africa, little evidence supports the idea that two lineages of chimpanzees are present in those forests. In fact, the findings of this study provide support for Goldberg's (1996) hypothesis that chimpanzees inhabiting forests in eastern Africa probably are the product of a recent and explosive expansion of chimpanzees, but probably from a core chimpanzee population that persisted in western equatorial Africa. The relationship between chimpanzees in Upper Guinea and those in Nigeria and western Cameroon appears to be more distant. Both populations probably were established earlier than populations in eastern Africa; and have been more isolated from each other and from chimpanzees in western equatorial African forests. Following the initial expansion of chimpanzees into the forests of western Africa, the two populations may have been semi-isolated, only sporadically exchanging migrants. However, it is not entirely clear whether the Dahomey Gap or the Niger River has been more important in defining the biogeographic relationship between these two regional populations.

The mtDNA and STR databases leave two taxonomic alternatives for chimpanzees. Given the consistency of the analyses of the mtDNA database, the most parsimonious solution is to divide chimpanzee populations into two subspecies that reflect the ancient western and central division: *P. t. verus* and *P. t. troglodytes* (Figure 5.2a). In this model, *P. t. verus* ranges from Upper Guinea to western Cameroon; and *P. t. troglodytes* from southern Cameroon to the eastern rim of African tropical forests.

The evidence of subdivision within those clades is less convincing. The results of this project do not support Morin *et al.*'s (1994) hypothesis that chimpanzees in Upper Guinea are sufficiently different from other chimpanzees to warrant full species status. The addition of more samples from Nigeria and Cameroon to the pan-African genetic database suggest also that chimpanzees in western Africa have had a more complex relationship than that proposed by Gonder *et al.* (1997). Despite the fact that chimpanzees in western equatorial and eastern Africa appear to have a relatively recent historical association, these two geographically defined populations are still usually considered different subspecies. Given that the two putative lineages in western Africa are more different from one another than the western equatorial and eastern African populations, then two forms of western African chimpanzee should probably be recognized if the current taxonomy of western equatorial and eastern African chimpanzees is maintained. In this model of chimpanzee evolution four subspecies would be recognized (figure 6.2b): Upper Guinea *P. t. verus*, western African *P. t.*

vellerosus, western equatorial African *P. t. troglodytes* and eastern African *P. t. schweinfurthii*.

Chimpanzee paleodemography and African forest history

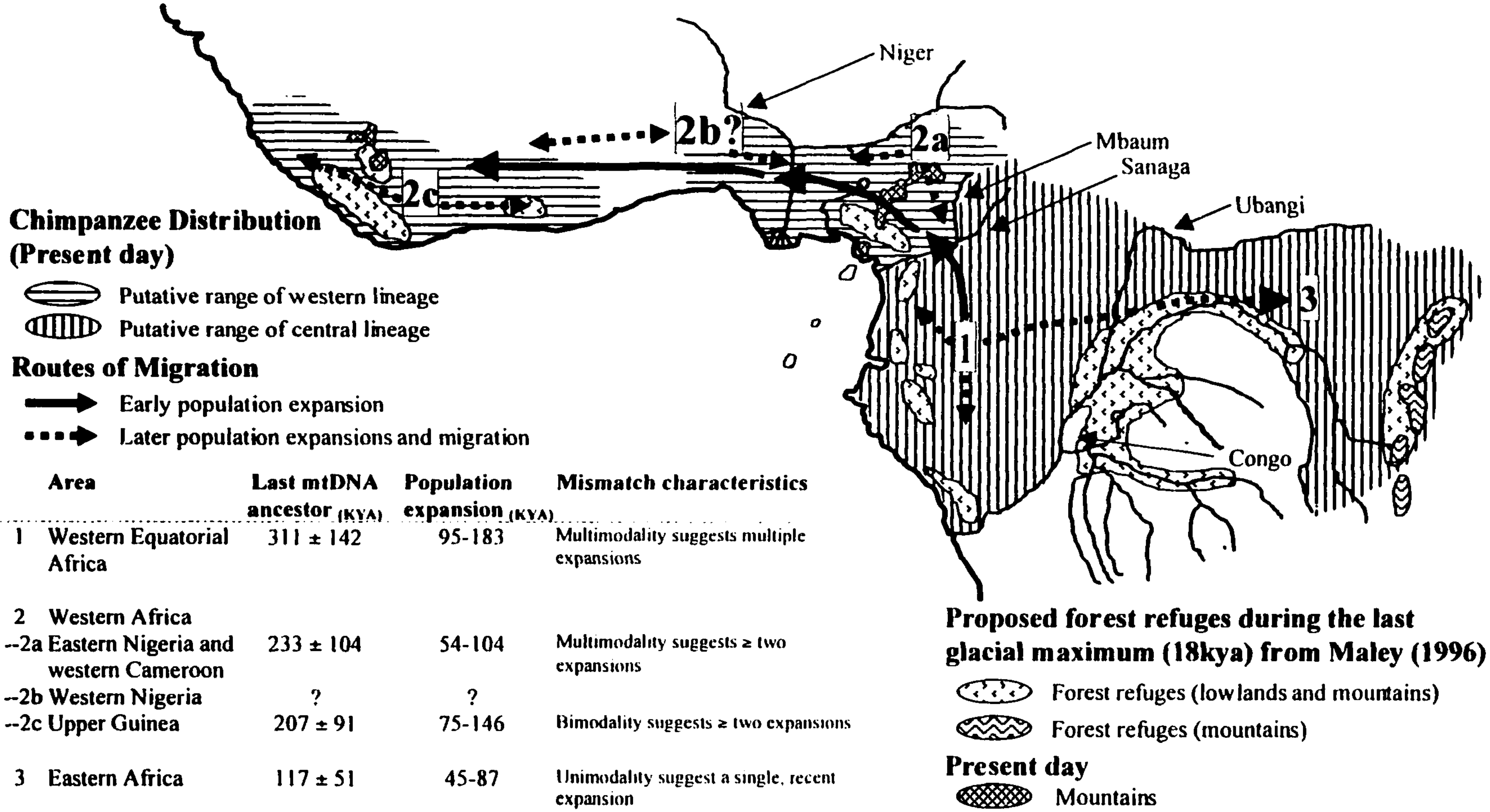
Chapter 1 detailed the putative relationship between the history of African forests and the recent evolution of forest-dependant species. Briefly, different lineages cluster into distinct communities that may reflect a shared historical relationship with the forests they inhabit. The Pleistocene Refuge hypothesis predicts that during periods of maximum glaciation, tropical rainforest animals were isolated into refuges brought about by the reduction and fragmentation of tropical forests. In this model, rainforest species differentiated due to the isolation of local populations in refuges and/or due to events accompanying the dispersal of these species during post-glacial forest expansion (Grubb 1982; Hamilton 1988; Grubb 1990). These forces have probably influenced the recent evolution and distribution of species over a series of glacial cycles spanning the Pleistocene (Livingstone 1993; Maley 1996).

The locations of Pleistocene refuges are often inferred based on an area's high level of species' endemism and diversity. High species' endemism and diversity suggest that populations within these forests persisted despite global shifts in forest size and composition during glacial episodes (Booth 1958a, 1958b; Grubb 1982; Haffer 1982; Mayr and O'Hara 1986; Hamilton 1988; Kingdon 1989; Grubb 1990; Colyn, Gautier-Hion, and Verheyen 1991; Hamilton 1992). Across western Africa, three main refuge locations have been proposed: one in Sierra Leone/Liberia, one in eastern Ivory Coast/Western Ghana and one in Cameroon/Gabon (Hamilton 1988). Given the complex distribution patterns of species within Nigeria and Cameroon, several smaller refuges may have persisted there during periods of maximum glaciation, for instance, in the Niger Delta (Moreau 1969) and the area between the Cross and Sanaga Rivers (Oates 1988; Maley 1996; Nichol 1999). These smaller refuges probably were probably separated from the larger core refuge area proposed south of the Sanaga River into Gabon (Maley 1996).

The concepts of endemism and diversity can be translated into genetic terms to make statements about the locations of refuges and of lineage expansions out of those areas that occurred as the climate ameliorated and the forests re-expanded. Assuming an island model of population genetic structure, samples from areas containing more "endemic" alleles, as well as, a more diverse allelic complement (i.e., sequences containing more pairwise nucleotide differences) may indicate the persistence of ancient populations (Harpending *et al.* 1993; Marjoram and Donnelly 1994; Rogers and Jorde 1995; Goldberg 1996, 1998). Populations with high levels of allelic endemism and diversity may have been protected within Pleistocene refuges during periods of maximum glaciation (Braemer 1992; Morrone and Crisci 1995; Goldberg 1996; Goldberg and Ruvolo 1997; Goldberg 1998). The distribution of genetic diversity among chimpanzees appears to follow that hypothesis. The pan-African mtDNA database suggests that global climatic fluctuations have influenced chimpanzee distribution and evolution possibly over a series of glacial cycles spanning the last few hundred thousand years. Although it is impractical to use the mtDNA database to precisely calibrate times of population range expansions with particular glacial cycles, analyses of this database reveal interesting parallels with respect to the location of Pleistocene refuges and periods of forest expansion.

Figure 6.3 depicts a model of chimpanzee migration and range expansions during the Pleistocene inferred from analyses of the HV-I sequence database. Of all chimpanzee mtDNA HV-I sequences, those of chimpanzees from western equatorial African forests are more "endemic" and diverse. For example, HV-I mtDNA sequences were shared only between individuals within sampling locations in southern Cameroon. In contrast, among chimpanzees in Upper Guinea, identical mtDNA sequences were shared between populations spanning 1,000 km (Gagneux 1998). Similarly, identical mtDNA HV-I sequences were shared between chimpanzee populations in eastern Africa spanning nearly 600 km (Goldberg 1996). Pairwise differences of mtDNA sequences between chimpanzees in western equatorial Africa revealed a higher average of nucleotide substitutions between sequences from samples in that region compared to any other region. Consequently, chimpanzees have probably persisted longer in western equatorial African forests than in any other area. These findings confirm the conclusions of other studies that suggest that the most genetically and phenotypically diverse of all chimpanzees are found in western equatorial Africa (Rothschild 1905; Rothschild 1906; Elliot 1913; Goodman *et al.* 1967; Hill 1967; Goodman and Tashian 1969; Hill 1969; Shea and Coolidge 1988; Groves, Westwood, and Shea 1993; Morin *et al.* 1994; Kaessmann, Wiebe, and Pääbo 1999).

The multimodal peaks and population expansion dates calculated from the intermatch and mismatch distributions presented in Chapter 4 imply that a core chimpanzee population persisted in western equatorial African forests over the last few



Area	Last mtDNA ancestor (KYA)	Population expansion (KYA)	Mismatch characteristics
1 Western Equatorial Africa	311 ± 142	95-183	Multimodality suggests multiple expansions
2 Western Africa			
--2a Eastern Nigeria and western Cameroon	233 ± 104	54-104	Multimodality suggests ≥ two expansions
--2b Western Nigeria	?	?	
--2c Upper Guinea	207 ± 91	75-146	Bimodality suggests ≥ two expansions
3 Eastern Africa	117 ± 51	45-87	Unimodality suggest a single, recent expansion

Figure 6.3. Relationship inferred between African forest history and chimpanzee paleodemography during the Pleistocene. This model was inferred from the analysis of mismatch distributions for the HV-I locus presented in Chapter 4. It predicts that chimpanzee populations expanded from forests in western equatorial Africa (1) into western African forests (2) during a mid-Pleistocene glacial episode. Following the initial population expansion (1 → 2), populations in western Africa (2a, b, c) have experienced population bottlenecks and re-expansions largely independent of chimpanzees in western equatorial Africa. Chimpanzees colonized forests in eastern Africa (3) most recently, and possibly at the same time as the latest re-expansion of chimpanzees into western Africa.

hundred thousand years, despite global climatic oscillations. However, the jaggedness of the western equatorial African mismatch distribution suggests that this population has not been completely impervious to global climatic shifts. While it is likely that this population persisted within a western equatorial African refuge, range contractions and expansions probably occurred periodically. These expansions and contractions may have followed the fragmentation and expansion of the forest cover over a series of glacial cycles, because these peaks of mismatch distributions correspond to demographic bottlenecks (Harpending 1994; Sherry *et al.* 1994; Rogers *et al.* 1996).

Chimpanzees in Upper Guinea and those in Nigeria and western Cameroon probably have had a more complex history, but may represent the next most ancient radiation of chimpanzees into African forests. Population expansion dates calculated from the mismatch distributions of chimpanzee HV-I sequences from these two regions imply a major population expansion of chimpanzees from western equatorial African forests into the forests of western Africa roughly 100 to 150kya. The similar population expansion dates but dissimilar shapes and modalities of their mismatch distributions (Marjoram and Donnelly 1994) suggest that chimpanzees in Upper Guinea and those in Nigeria and western Cameroon may represent two semi-isolated parts of a fairly ancient radiation of chimpanzees into western Africa.

Moreover, the dissimilar shapes of the mismatch distributions of HV-I sequences of chimpanzees from Upper Guinea and those in Nigeria and western Cameroon, suggest that these populations have experienced separate paleodemographic events and probably persisted in two separate refuges during subsequent climatic deterioration. Chimpanzees in Upper Guinea may have persisted in a Liberian refuge (Gagneux 1998), following the initial expansion of this species into the forests of western Africa. Given that identical sequences were detected in populations spanning 1,000 km. and the large smooth wave of the mismatch for this population, it seems likely that there was a more recent expansion of chimpanzees from a refuge in Liberia into newly re-expanded chimpanzee habitat. This population expansion may have occurred after the last glaciation, or perhaps after an earlier glacial episode.

During a period of aridity following population expansions around 100-150kya, the eastern Nigeria and western Cameroon population possibly persisted in a small refuge between the Cross and Sanaga Rivers, perhaps in the vicinity of the Cameroon Highlands. Given the multimodal peaks of the Nigeria-western Cameroon mismatch distribution, it seems likely that this population has undergone smaller and more frequent demographic bottlenecks and population expansions than chimpanzee populations in Upper Guinea. Moreover, the persistence of the leading waves of the Nigeria-Cameroon intermatch distribution in the eastern Nigeria-western Cameroon mismatch distribution implies that some ancestral sequences are still present in that population. Those persistent peaks could be the product of recent individual migrations across the Sanaga River (i.e., within the last few generations) or they could be signatures of more recent expansions of chimpanzees from western equatorial Africa that followed the major expansion of this species throughout the forests of western Africa.

Chimpanzees probably expanded most recently into the forests of eastern Africa. The calculated population expansion dates for chimpanzees in eastern Africa suggest that this event occurred roughly 75kya, or perhaps more recently. The unimodal mismatch distribution for this population suggests that, unlike other regional chimpanzee populations, chimpanzees in eastern Africa have not undergone semi-isolated range contractions and expansions (Harpending *et al.* 1993; Marjoram and Donnelly 1994). In addition, two "eastern African" HV-I sequences were detected in the central Cameroon region. Those sequences could be the result of recent migration, but it is also possible that those two sequences represent the last local representatives of nearly extinct "eastern African" HV-I haplotypes that predate the widespread expansion of chimpanzees throughout eastern forests.

Goldberg (1996, 1998) proposed that chimpanzees persisted in eastern Africa during periods of maximum glaciation by exploiting marginal savanna-like habitat. In his model, after the amelioration of climate and the subsequent re-expansion of the rainforest, chimpanzees recolonized their former habitat. Due to the increased carrying capacity of their recovered habitat, chimpanzee populations grew explosively. The analyses presented in this dissertation do not support that view. The close relationship between chimpanzees in western equatorial and eastern Africa, the lack of ancient eastern African mtDNA sequences that might predate a massive expansion of chimpanzees into eastern forests and the presence of "eastern African" mtDNA sequences in Cameroon suggest that another paleodemographic scenario is likely.

It is possible that prior to 75kya, chimpanzees in western equatorial and eastern Africa were isolated in a single refuge, perhaps a large refuge in southern Cameroon and Gabon. Following the expansion of the rainforest into eastern Africa, it is possible that chimpanzees migrated rapidly from that refuge eastward into the newly re-expanded rainforest. Given the fairly continuous forest coverage from western equatorial to eastern Africa, a recent population expansion combined with high levels of gene flow may have maintained the similarities between these two regional populations. This scenario is more parsimonious that Goldberg's (1996, 1998) hypothesis because it does not require fundamental shifts in this species' ecology to account for the observed apportionment of genetic diversity. However, there are virtually no samples of chimpanzees from interior portions of western equatorial Africa, such as, Congo, Central African Republic and the northern and western portions of the Democratic Republic of Congo. Until samples are obtained from those areas, this particular hypothesis cannot be properly evaluated.

It is interesting, that despite the dissimilarities in the mismatch distributions of each regional chimpanzee population, each mismatch contained a peak within the 12-20 base pair range. This peak may correspond to population expansions about 65 to 75kya. Given these similarities, it is possible that at the same time chimpanzees were expanding into eastern Africa, there may have been smaller expansions of the more ancient western and western equatorial African populations. However, this hypothesis requires a secondary population expansion of chimpanzees into western African forests.

There may have been multiple population expansions of chimpanzees that tracked the series of cool, arid advances during the Upper Pleistocene. Although the last period of maximum glaciation is proposed to have been the most extreme (Maley 1996), some evidence suggests that an earlier mid-Pleistocene glacial advance may have nearly destroyed the rainforest in western Africa. For instance, Nichol (1999) analyzed the geomorphology of fossil sand dunes in central Nigeria (see Chapter 1). Based on her analyses of these fossil sand dunes she suggested that the most extreme glacial episode occurred during the mid-Pleistocene, either 250, 190-155 or 120-90 kya. Based on the southerly extent of the fossil sand dunes, she suggested that during this episode nearly all of the forests across western Africa may have been destroyed. This mid-Pleistocene glacial advance may have affected chimpanzees across western Africa. Chimpanzee populations could have been all but eliminated during that period. After the amelioration of climate, they may have recolonized their former habitats. Although this hypothesis predicts that a secondary expansion of chimpanzees into western Africa occurred during the mid-Pleistocene, later population expansions are also possible (i.e., after the last glaciation).

Overall, the mtDNA database suggests that over a series of Pleistocene glacial episodes a core chimpanzee population may have persisted in a western equatorial African forest refuge. From that core population, chimpanzees may have expanded into western Africa about 100 to 150kya. Subsequent to that expansion, chimpanzees in Upper Guinea and those in Nigeria and western Cameroon may have been semi-isolated from one another. Chimpanzee populations may have expanded most recently into eastern Africa, perhaps beginning 65kya. These chimpanzees may have expanded from a large western equatorial African forest refuge, and not from marginal savanna-like or riparian habitats. At the same time that chimpanzees expanded into eastern forests, the mtDNA data suggest that the chimpanzee populations in western Africa also may have experienced smaller population expansions at approximately the same time.

Future work

There are ample opportunities for future research on this topic. The results of this project suggest changes to the widely accepted views of recent chimpanzee evolution and taxonomy, but these changes are based on information from only small portions of two genomes. Much more information can be gleaned from the existing sample base. Several researchers have published nuclear DNA sequences that have yielded conflicting information about intraspecific variation among the apes (e.g., Deinard and Kidd 1996; Deinard *et al.* 1999; Deinard and Kidd 1999, 2000). For example, in a recent study Kaesmann, Wiebe, and Pääbo (1999), analyzed intraspecific variation among chimpanzees at the Xq3.13 locus of the X chromosome. They found no separation of the widely accepted chimpanzee subspecies. The Xq3.13 data might conflict with the mtDNA database for several reasons, for instance, a slower mutation rate or the three fold increase in effective population size of the X chromosome. These factors could make variation at the Xq3.13 locus predate the separation of the subspecies (Kaessmann, Wiebe, and Pääbo 1999). Moreover, their sample size may have been too small to discriminate the differences between chimpanzee lineages with any precision. The addition of more samples of chimpanzees at that locus may more clearly resolve the apportionment of genetic diversity between different lineages. Consequently, data from other genomic regions, including the Xq3.13, should be evaluated to look for other possible scenarios of recent chimpanzee evolution and for other phylogeographic divisions of chimpanzee populations.

The Sanaga River is the first phylogeographic break to be identified using genetic data for any African ape and provides an unparalleled opportunity to examine recent ape evolution from many different perspectives. A first step towards achieving that goal would be a more intensive survey of chimpanzee populations in central and eastern Cameroon. For example, more extensive sampling near Manb'ra village (map location, 13) may more clearly resolve how extensive gene flow has been between the western and central African chimpanzee lineages. More research is also needed into what maintains this deep genetic division between chimpanzees in the region. The Sanaga River is not large, and there is no direct evidence that the course and size of the river has changed significantly during the Pleistocene (J. Nichol, pers. comm.). It seems likely that other factors also maintain this deep genetic break. Behavioral and ecological differences are known to exist between chimpanzee lineages (Whiten *et al.* 1999). More intensive fieldwork might reveal that behavioral and/or ecological differences between western and central African chimpanzees augments the influence of the Sanaga River in limiting gene flow between lineages.

The Sanaga River probably has had a considerable effect on the evolution of many primates and other mammals, but it is unlikely that this river is the only boundary producing the high biodiversity in present-day Nigeria and Cameroon. Many geographic features in this region of Africa, like the Dahomey Gap and the Niger and Cross Rivers, may have exerted considerable influence on the distribution of many species, and probably on very different time scales (Booth 1958a, 1958b; Moreau 1969; Grubb 1982; White 1983; Grubb 1990; Groves 1993; Maley 1996). Consequently, a more general genetic survey of the phylogeography of primates and other mammals is needed to explore the influence of the Sanaga River and of other biogeographic boundaries in the vicinity of Nigeria and Cameroon.

Sampling gaps remain in the pan-African database that make the existing interpretation of the data tenuous. The Niger River may play a prominent role in recent chimpanzee evolution, but the sample size of chimpanzees from western Nigeria was too small to allow this hypothesis to be directly tested. During the sampling phase of this project, several local informants provided reliable information that chimpanzees were present in the Ala, Owo and Ogbesse Forest Reserves that are south and east of the Ise Forest Reserve. Unfortunately, time constraints prevented a full investigation of these reports. Consequently, more extensive efforts should be made to sample from and learn more about these populations. Moreover, the "central African" HV-I sequences found in the Cross River National Park are intriguing. It is possible that other biogeographic barriers, like the Cross River, have augmented the influence of the Sanaga River. Given these possibilities, more extensive sampling in southeastern Nigeria is warranted.

There are sampling gaps in other regions of Africa. Most prominent among them is the virtual lack of samples from interior portions of western equatorial Africa. Given that chimpanzees in western equatorial and eastern Africa are not genetically well-differentiated from one another, the addition of samples from this intervening area might clarify the distribution of genetic diversity among chimpanzee populations across the Congo River basin. Finally, in order to tease apart the respective influences of the Niger River and the Dahomey Gap, more extensive sampling along the perimeter of the Gap may prove enlightening. Chimpanzees still survive in western Ghana (Oates 1996; Oates, Struhsaker and Whitesides 1997; M. Abedi, pers. comm. to J. Oates). Although western Ghana is not very close to the Dahomey Gap, any samples that could be obtained from chimpanzees there may prove useful in evaluating the importance of the Gap in recent chimpanzee evolution.

Chimpanzee conservation

Chimpanzees are probably extinct, or nearly so, in about half the countries where they have historically ranged (Lee, Thornback, and Bennett 1988; Teleki 1989; Oates 1996). Many factors have contributed to their decline. Vast tracts of suitable chimpanzee habitat have been lost or fragmented due to slash and burn agriculture, plantation monocropping and timber harvesting (Western and Pearl 1989; Sayer, Harcourt, and Collins 1992; Harcourt 1996). Chimpanzees are often hunted for food, trophies, and historically in western Africa, for the biomedical industry (Sayer, Harcourt, and Collins 1992; Oates 1996; Prescott, Rapley, and Joseph 1996; Williamson and Usongo 1996). Even in areas where they are protected, poaching is very common (Struhsaker 1987; Teleki 1989). Consequently, even in habitats where chimpanzees can survive there has been a precipitous decline of the remaining populations.

Precise population size estimates are very difficult to calculate. They are often based on very imprecise information such as, percentage of forest cover, rates of forest disappearance, human population density, gross national products and animal population densities that are calculated from well-studied populations and forests that receive some degree of protection (Meffe and Carroll 1994; Harcourt 1996). Given these caveats, western equatorial African forests may shelter the largest remaining chimpanzee populations. A minimum of 80,000 chimpanzees may survive in Gabonese and Congolese forests (Teleki 1989; Oates 1996). The situation in western Africa is much less encouraging. As few as 12,000 chimpanzees may survive in western Africa (Lee, Thornback, and Bennett 1988; Oates 1996), but there may be many more (J. F. Oates, pers. comm.).

Prior to this study, chimpanzees were widely believed to be extinct in Nigeria and rare in Cameroon (Goodall 1986; Teleki 1989; Morin *et al.* 1994). While the field-survey component of this project proves that chimpanzees still survive in the area, there is little cause for optimism. In Nigeria and Cameroon, chimpanzees face a serious extinction threat. Widespread forest clearance.and bushmeat hunting occurs almost completely unchecked (Gadsby and Jenkins 1992; Oates 1996; Prescott, Rapley, and Joseph 1996; Williamson and Usongo 1996). In the coming decades, it is likely that chimpanzees will survive only in a few fragmented pockets of forest. Of those, it is very likely that only heavily protected areas will ensure the survival of their resident populations. Even with that protection, the survival of this species is still threatened.

It is unwise to prioritize conservation efforts on the basis of genetics alone. However, from a genetics point of view, there are two conservation issues to be faced: the consequences of habitat fragmentation and the identification of conservation units within the species. Habitat fragmentation is widely acknowledged to pose serious threats to a species' long-term survival (Soulé 1986, 1987; Woodruff 1990; Meffe and Carroll 1994; Avise and Hamrick 1996). Isolated local populations are more prone to extinction than large populations as a result of stochastic demographic factors like resource depletion, unequal sex ratios and the spread of infectious diseases (Gilpin and Soulé 1986; Avise and Hamrick 1996). Over the long-term, local population isolation can have deleterious genetic effects. These effects initially include a loss of genetic diversity, and eventually, inbreeding and extinction due to the potential for exposing recessive lethal alleles (Woodruff 1990; Meffe and Carroll 1994; O'Brien 1994). Populations of some species already exhibit signs of the long term effects of genetic isolation that are attributed to habitat fragmentation (Wildt *et al.* 1987; Laurenson and Caro 1994; Avise and Hamrick 1996).

At present, gene flow among chimpanzee populations appears to have been extensive within lineages. Among populations that have been extensively sampled, only chimpanzees at Taï in Ivory Coast and at Gombe in Tanzania have been reported to show signs of local population isolation and a loss of genetic diversity that could have resulted from habitat fragmentation (Morin 1992; Gagneux 1998). In most areas, chimpanzees in Nigeria and Cameroon exist in even more isolated and fragmented populations, and very few chimpanzees probably remain in western Nigeria. Nevertheless, there is little direct evidence to support the idea that chimpanzees in the area have suffered a loss of genetic diversity that is attributable to local population isolation. Given the level of widespread forest clearance and high hunting pressure, the coming decades may witness large population declines, the widespread extinction of isolated local populations and possibly a significant loss of genetic diversity.

Conservation biologists often acknowledge that corridors may reduce the deleterious effects of habitat fragmentation (Soulé 1986; Fahrig and Merriam 1994; Meffe and Carroll 1994). Maintaining corridors has often been promoted in reserve design to allow movement between forest remnants. On a broader scale, corridors can allow for the exchange of migrants between populations to maintain the genetic cohesiveness of different populations and lineages (Woodruff 1990; Meffe and Carroll 1994; Ayala and Escalante 1996). One problem in determining the best location for a corridor is that it is often unclear where migrants are commonly exchanged (Fahrig and Merriam 1994).

Chimpanzees of different lineages do migrate between populations in eastern Cameroon. Analyses of the mtDNA sequences from chimpanzees east of the Mbaum River suggested that migration between the western and central African lineages is mostly limited to that area. Given the potential significance of that region as a gene flow corridor between chimpanzee lineages, more extensive conservation efforts should be concentrated there. Eastern Cameroon appears to be an important migration corridor for other species as well (B. Curran and T. Smith, unpublished data). Consequently, focusing on this area may provide additional benefits to other species.

Genetic data are often used by conservation biologists to identify evolutionarily significant units, or ESU's. An ESU is defined as a population that is reproductively isolated from other conspecific population units, and that represents an important component in the evolutionary legacy of the species (Meffe and Carroll 1994; Avise and Hamrick 1996). Several researchers have proposed different methods to rank-order the importance of different populations (e.g., Avise 1989; O'Brien and Mayr 1991; Templeton 1991; Baker 1994; Faith 1994; Giles 1994; Templeton 1994). Despite the serious extinction threats chimpanzees face, it is unwise to make recommendations for the conservation of any population on the basis of genetic significance or uniqueness alone. If certain populations are decidedly not unique or are considered evolutionarily insignificant, that type of paradigm could easily result in the neglect and eventually the destruction or extinction of whole populations or even lineages.

The debate over the phylogenetic and taxonomic status of chimpanzees will continue. The relative "importance" of different populations may change as more information is gathered. In addition, recent chimpanzee evolutionary history may never be resolved to an indisputable level of scientific certainty. Nevertheless, it is possible to set certain conservation priorities for chimpanzees in western Africa. Given the intriguing evidence that chimpanzees in western Africa might be divided into two separate lineages, it is prudent at this point to recognize chimpanzees in Upper Guinea and those in Nigeria and western Cameroon as independent evolutionary units within the species for conservation purposes.

Moreover, given the potential significance of the Niger River in separating these two putative lineages, it would be wise to focus more intensive conservation efforts on behalf of chimpanzees in western Nigeria. Although it will be difficult to prevent the eventual extinction of chimpanzees in western Nigeria, several forest reserves near Akure may harbor a few isolated populations (Agbelusi 1994; per. obs.). Virtually nothing is known about these populations, yet nothing is being done to even assess the size of the remaining populations. An intensive survey of the area might reveal that the conservation outlook for chimpanzees in western Nigeria might not be as bleak as the field-survey of this project suggested. In addition, a more intensive survey of chimpanzees in western Nigeria may reveal potential conservation strategies to preserve these populations such as, maintaining or creating migration corridors between populations. This survey would also have other benefits including, increasing publicity against hunting the remaining populations and the continued destruction of the reserves and conservation areas in western Nigeria.

Conserving chimpanzees across Africa is a daunting task. In Nigeria and Cameroon that task may be even more difficult due to a dense human population, particularly in Nigeria. Yet despite discouraging reports suggesting the near-extinction of chimpanzees in Nigeria and Cameroon, this project demonstrates that some large populations still persist in the region. The predecessors of these communities played a pivotal role in the recent evolution of this species. In the future, these chimpanzee communities may be very important again in maintaining the diversity present among wild chimpanzees. Given the biological significance of chimpanzees in Nigeria and Cameroon, every effort should be made to ensure their continued survival.

APPENDIX 1

Sample Catalogue

Map	Sample	Location	Approximate Coordinates	Hairs	Nest Age
1	ISEFR1	ISE FOREST RESERVE--Ekiti State	N 07°22.103' E 05°21.869'	9	2
1	ISEFR2	ISE FOREST RESERVE--Ekiti State	N 07°22.103' E 05°21.869'	3	2
1	ISEFR3	ISE FOREST RESERVE--Ekiti State	N 07°22.103' E 05°21.869'	30	3
1	ISEFR4	ISE FOREST RESERVE--Ekiti State	N 07°22.103' E 05°21.869'	18	2
1	ISEFR5	ISE FOREST RESERVE--Ekiti State	N 07°22.103' E 05°21.869'	7	1
2	OWOFR1	OWO FOREST RESERVE--Ondo State	N 06°58.399' E 05°32.136'	2	5
2	OWOFRRC1	OWO FOREST RESERVE--Ondo State	N 06°59.071' E 05°31.522'	10	N/A
2	OWOFRRC2	OWO FOREST RESERVE--Ondo State	N 06°59.071' E 05°31.522'	35	N/A
2	OWOFRRC3	OWO FOREST RESERVE--Ondo State	N 06°59.071' E 05°31.522'	7	N/A
2	OWOFRLH1	OWO FOREST RESERVE--Ondo State	N 06°58.399' E 05°32.136'	1	N/A
2	OWOFRLH2	OWO FOREST RESERVE--Ondo State	N 06°59.071' E 05°31.522'	1	N/A
2	OWOFRLH3	OWO FOREST RESERVE--Ondo State	N 06°59.071' E 05°31.522'	1	N/A
2	OWOFRLH4	OWO FOREST RESERVE--Ondo State	N 06°58.399' E 05°32.136'	1	N/A
3	CRNP1	CROSS RIVER NATIONAL PARK--Oban Hills	≈ N 07°28' E 08°10'	2	2
3	CRNP2	CROSS RIVER NATIONAL PARK--Oban Hills	≈ N 07°28' E 08°10'	30	2
3	CRNP3	CROSS RIVER NATIONAL PARK--Oban Hills	≈ N 07°28' E 08°10'	14	3
3	CRNP4	CROSS RIVER NATIONAL PARK--Oban Hills	≈ N 07°28' E 08°10'	7	3
3	CRNP5	CROSS RIVER NATIONAL PARK--Oban Hills	≈ N 07°28' E 08°10'	7	2
3	CRNP6	CROSS RIVER NATIONAL PARK--Oban Hills	≈ N 07°28' E 08°10'	13	2
3	CRNP7	CROSS RIVER NATIONAL PARK--Oban Hills	≈ N 07°28' E 08°10'	6	2
3	CRNP8	CROSS RIVER NATIONAL PARK--Oban Hills	≈ N 07°28' E 08°10'	25	2
3	CRNP8	CROSS RIVER NATIONAL PARK--Oban Hills	≈ N 07°28' E 08°10'	10	2

Map	Sample	Location	Approximate Coordinates	Hairs	Nest Age
3	CRNP10	CROSS RIVER NATIONAL PARK--Oban Hills	≈ N 07°28' E 08°10'	12	2
4	CRNP11	CROSS RIVER NATIONAL PARK--Okwangwo	N 06°23.8' E 9°19.6'	17	2
4	CRNP12	CROSS RIVER NATIONAL PARK--Okwangwo	N 06°23.8' E 9°19.6'	15	2
5	AKZN1	AKOH ZANTO	N 06°54.644' E 10°53.993'	3	2
5	AKZN2	AKOH ZANTO	N 06°54.644' E 10°53.993'	15	2
5	AKZN3	AKOH ZANTO	N 06°54.644' E 10°53.993'	50	2
5	AKZN4	AKOH ZANTO	N 06°54.644' E 10°53.993'	4	2
5	AKZN5	AKOH ZANTO	N 06°54.644' E 10°53.993'	7	2
5	AKZN6	AKOH ZANTO	N 06°54.644' E 10°53.993'	15	2
5	AKZN7	AKOH ZANTO	N 06°54.644' E 10°53.993'	17	2
5	AKZN8	AKOH ZANTO	N 06°54.644' E 10°53.993'	25	2
5	AKZN9	AKOH ZANTO	N 06°54.644' E 10°53.993'	18	2
5	AKZN10	AKOH ZANTO	N 06°54.644' E 10°53.993'	23	2
5	AKZN11	AKOH ZANTO	N 06°54.644' E 10°53.993'	13	3
5	AKZN12	AKOH ZANTO	N 06°54.644' E 10°53.993'	32	2
5	AKZN13	AKOH ZANTO	N 06°54.644' E 10°53.993'	3	2
5	AKZN14	AKOH ZANTO	N 06°54.644' E 10°53.993'	48	2
5	AKZN15	AKOH ZANTO	N 06°54.644' E 10°53.993'	14	2
5	AKZN16	AKOH ZANTO	N 06°54.644' E 10°53.993'	36	2
6	NGNY1	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	10	2
6	NGNY2	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	12	2
6	NGNY3	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	3	3
6	NGNY4	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	5	3
6	NGNY5	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	4	4
6	NGNY6	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	2	4
6	NGNY7	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	4	3
6	NGNY8	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	5	3
6	NGNY9	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	18	1
6	NGNY10	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	10	1
6	NGNY11	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	12	3
6	NGNY12	NGEL NYAKI FOREST RESERVE	N 07°04.412' E 11°02.810'	5	3

Map	Sample	Location	Approximate Coordinates	Hairs	Nest Age
6	NGNY13	NGEL NYAKI FOREST RESERVE	N 07°04.412' E 11°02.810'	6	3
6	NGNY14	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	20	2
6	NGNY15	NGEL NYAKI FOREST RESERVE	≈N 07°04' E 11°04'	400	N/A
8	GGNP1	GASHAKA-GUMTI NATIONAL PARK--Kwano	≈N 07°19.73' E 11°35.64'	2	2
8	GGNP2	GASHAKA-GUMTI NATIONAL PARK--Kwano	≈N 07°19.73' E 11°35.64'	6	2
8	GGNP3	GASHAKA-GUMTI NATIONAL PARK--Kwano	N 07°19.733' E 11°35.644'	10	3
8	GGNP4	GASHAKA-GUMTI NATIONAL PARK--Kwano	N 07°19.733' E 11°35.644'	11	3
8	GGNP5	GASHAKA-GUMTI NATIONAL PARK--Kwano	N 07°19.733' E 11°35.644'	6	1
8	GGNP6	GASHAKA-GUMTI NATIONAL PARK--Kwano	N 07°19.733' E 11°35.644'	8	1
8	GGNP7	GASHAKA-GUMTI NATIONAL PARK--Kwano	N 07°19.733' E 11°35.644'	10	2
8	GGNP8	GASHAKA-GUMTI NATIONAL PARK--Kwano	N 07°19.733' E 11°35.644'	20	2
8	GGNP9	GASHAKA-GUMTI NATIONAL PARK--Kwano	N 07°19.733' E 11°35.644'	11	3
8	GGNP10	GASHAKA-GUMTI NATIONAL PARK--Kwano	N 07°19.733' E 11°35.644'	12	2
8	GGNP11	GASHAKA-GUMTI NATIONAL PARK--Kwano	N 07°19.733' E 11°35.644'	23	2
8	GGNP12	GASHAKA-GUMTI NATIONAL PARK--Kwano	≈N 07°19.73' E 11°35.64'	12	1
8	GGNP13	GASHAKA-GUMTI NATIONAL PARK--Kwano	≈N 07°19.73' E 11°35.64'	2	1
8	GGNP14	GASHAKA-GUMTI NATIONAL PARK--Kwano	≈N 07°19.73' E 11°35.64'	4	2
8	GGNP15	GASHAKA-GUMTI NATIONAL PARK--Kwano	≈N 07°19.73' E 11°35.64'	5	4
8	GGNP16	GASHAKA-GUMTI NATIONAL PARK--Kwano	≈N 07°19.73' E 11°35.64'	15	3
8	GGNP17	GASHAKA-GUMTI NATIONAL PARK--Kwano	≈N 07°19.73' E 11°35.64'	2	3
8	GGNP18	GASHAKA-GUMTI NATIONAL PARK--Gumti	≈ N 07°38.24' E11°44.53'	13	3
9	GGNP19	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	15	2
9	GGNP20	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	11	2
9	GGNP21	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	14	2
9	GGNP22	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	6	2
9	GGNP23	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	30	2
9	GGNP24	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	19	2
9	GGNP25	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	7	2
9	GGNP26	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	15	2
9	GGNP27	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	18	2
9	GGNP28	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	20	2

Map	Sample	Location	Approximate Coordinates	Hairs	Nest Age
9	GGNP29	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	4	2
9	GGNP30	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	7	2
9	GGNP31	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	11	2
9	GGNP32	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	25	2
9	GGNP33	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	15	2
9	GGNP34	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	17	2
9	GGNP35	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	24	2
9	GGNP36	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	3	2
9	GGNP37	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	5	2
9	GGNP38	GASHAKA-GUMTI NATIONAL PARK--Gumti	N 07°38.243' E11°44.534'	11	2
9	GGNP39	GASHAKA-GUMTI NATIONAL PARK--Leinde Fadali	N 06°59.009' E 11°36.833'	8	?
7	GGNP40	GASHAKA-GUMTI NATIONAL PARK--Leinde Fadali	N 06°59.009' E 11°36.833'	20	?
7	GGNP41	GASHAKA-GUMTI NATIONAL PARK--Leinde Fadali	N 06°59.724' E 11°36.465'	18	?
7	GGNP42	GASHAKA-GUMTI NATIONAL PARK--Leinde Fadali	N 06°59.724' E 11°36.465'	2	?
7	GGNP43	GASHAKA-GUMTI NATIONAL PARK--Leinde Fadali	N 06°59.724' E 11°36.465'	15	?
7	GGNP44	GASHAKA-GUMTI NATIONAL PARK--Leinde Fadali	N 06°59.933' E 11°36.615'	2	?
7	GGNP45	GASHAKA-GUMTI NATIONAL PARK--Leinde Fadali	N 06°59.933' E 11°36.615'	7	?
7	GGNP46	GASHAKA-GUMTI NATIONAL PARK--Leinde Fadali	N 06°59.933' E 11°36.615'	7	?
10	BYM1	BANYANG-MBO RESERVE	N 05°19.829' E 09°24.904'	16	2
10	BYM2	BANYANG-MBO RESERVE	N 05°19.829' E 09°24.904'	1	2
10	BYM3	BANYANG-MBO RESERVE	N 05°19.829' E 09°24.904'	3	3
10	BYM4	BANYANG-MBO RESERVE	N 05°19.829' E 09°24.904'	22	4
10	BYM5	BANYANG-MBO RESERVE	N 05°19.829' E 09°24.904'	12	2
10	BYM6	BANYANG-MBO RESERVE	N 05°19.829' E 09°24.904'	15	4
11	MTCM1	MT. CAMEROON/MAPANJA	N 04°07.362' E 09°09.453'	20	2
11	MTCM2	MT. CAMEROON/MAPANJA	N 04°07.362' E 09°09.453'	5	2
11	MTCM3	MT. CAMEROON/MAPANJA	N 04°07.362' E 09°09.453'	20	5
11	MTCM4	MT. CAMEROON/MAPANJA	N 04°07.362' E 09°09.453'	40	1
11	MTCM5	MT. CAMEROON/MAPANJA	N 04°07.362' E 09°09.453'	15	1
11	MTCM6	MT. CAMEROON/MAPANJA	N 04°07.362' E 09°09.453'	21	1
11	MTCM7	MT. CAMEROON/MAPANJA	N 04°07.362' E 09°09.453'	20	1

Map	Sample	Location	Approximate Coordinates	Hairs	Nest Age
11	MTCM8	MT. CAMEROON/MAPANJA	N 04°07.362' E 09°09.453'	15	1
11	MTCM9	MT. CAMEROON/MAPANJA	N 04°07.362' E 09°09.453'	12	1
11	MTCM10	MT. CAMEROON/MAPANJA	N 04°07.362' E 09°09.453'	73	1
11	MTCM11	MT. CAMEROON/MAPANJA	N 04°07.362' E 09°09.453'	50	1
11	MTCM12	MT. CAMEROON/MAPANJA	N 04°07.515' E 09°08.626'	52	2
11	MTCM13	MT. CAMEROON/MAPANJA	N 04°07.515' E 09°08.626'	3	2
11	MTCM14	MT. CAMEROON/MAPANJA	N 04°07.515' E 09°08.626'	3	1
11	MTCM15	MT. CAMEROON/MAPANJA	N 04°07.515' E 09°08.626'	11	1
11	MTCM16	MT. CAMEROON/MAPANJA	N 04°07.515' E 09°08.626'	10	1
11	MTCM17	MT. CAMEROON/MAPANJA	N 04°07.515' E 09°08.626'	10	2
11	MTCM18	MT. CAMEROON/MAPANJA	N 04°07.515' E 09°08.626'	10	2
11	MTCM19	MT. CAMEROON/MAPANJA	N 04°07.515' E 09°08.626'	5	2
11	MTCM20	MT. CAMEROON/MAPANJA	N 04°07.515' E 09°08.626'	30	1
11	MTCM21	MT. CAMEROON/MAPANJA	N 04°07.515' E 09°08.626'	36	1
11	MTCM22	MT. CAMEROON/MAPANJA	N 04°07.515' E 09°08.626'	5	1
11	MTCM23	MT. CAMEROON/MAPANJA	N 04°07.515' E 09°08.626'	7	1
12	MOSSE1	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	3	1
12	MOSSE2	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	16	1
12	MOSSE3	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	3	4
12	MOSSE4	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	12	4
12	MOSSE5	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	16	3
12	MOSSE6	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	4	4
12	MOSSE7	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	1	1
12	MOSSE8	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	22	1
12	MOSSE9	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	8	1
12	MOSSE10	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	1	1
12	MOSSE11	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	21	1
12	MOSSE12	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	7	2
12	MOSSE13	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	14	2
12	MOSSE14	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	30	2
12	MOSSE15	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	29	2

Map	Sample	Location	Approximate Coordinates	Hairs	Nest Age
12	MOSSE16	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	10	2
12	MOSSE17	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	4	2
12	MOSSE18	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	7	2
12	MOSSE19	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	1	1
12	MOSSE20	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	5	1
12	MOSSE21	MOSSE/NORTH SANAGA	N 04°27.371' E 10°27.130'	15	1
13	MAN1	MANB'RA/NGAMBE TIKAR	N 05°37.281' E 11°40.855'	20	3
13	MAN2	MANB'RA/NGAMBE TIKAR	N 05°37.281' E 11°40.855'	20	5
13	MAN3	MANB'RA/NGAMBE TIKAR	N 05°37.281' E 11°40.855'	32	5
13	MAN4	MANB'RA/NGAMBE TIKAR	N 05°37.281' E 11°40.855'	6	5
13	MAN5	MANB'RA/NGAMBE TIKAR	N 05°37.281' E 11°40.855'	5	5
13	MAN6	MANB'RA/NGAMBE TIKAR	N 05°37.281' E 11°40.855'	23	2
13	MAN7	MANB'RA/NGAMBE TIKAR	N 05°37.281' E 11°40.855'	18	5
13	MAN8	MANB'RA/NGAMBE TIKAR	N 05°37.281' E 11°40.855'	19	3
13	MAN9	MANB'RA/NGAMBE TIKAR	N 05°37.281' E 11°40.855'	1	3
13	MAN10	MANB'RA/NGAMBE TIKAR	N 05°37.281' E 11°40.855'	3	3
13	MAN11	MANB'RA/NGAMBE TIKAR	N 05°37.281' E 11°40.855'	5	3
13	MAN12	MANB'RA/NGAMBE TIKAR	N 05°35.936' E11°43.828'	22	2
13	MAN13	MANB'RA/NGAMBE TIKAR	N 05°35.936' E11°43.828'	62	1
13	MAN14	MANB'RA/NGAMBE TIKAR	N 05°35.936' E11°43.828'	10	1
13	MAN15	MANB'RA/NGAMBE TIKAR	N 05°35.936' E11°43.828'	10	2
13	MAN16	MANB'RA/NGAMBE TIKAR	N 05°35.936' E11°43.828'	9	2
14	DEFR1	DOUALA-EDEA/TISSONGO	N 03°35.820' E 09°53.649'	12	2
14	DEFR2	DOUALA-EDEA/TISSONGO	N 03°35.820' E 09°53.649'	6	3
14	DEFR3	DOUALA-EDEA/TISSONGO	N 03°35.820' E 09°53.649'	9	2
14	DEFR4	DOUALA-EDEA/TISSONGO	N 03°35.820' E 09°53.649'	2	1
14	DEFR5	DOUALA-EDEA/TISSONGO	N 03°35.820' E 09°53.649'	1	1
14	DEFR6	DOUALA-EDEA/TISSONGO	N 03°35.820' E 09°53.649'	16	1
14	DEFR7	DOUALA-EDEA/TISSONGO	N 03°35.820' E 09°53.649'	7	1
14	DEFR8	DOUALA-EDEA/TISSONGO	N 03°35.820' E 09°53.649'	26	1
14	DEFR9	DOUALA-EDEA/TISSONGO	N 03°35.820' E 09°53.649'	3	1

Map	Sample	Location	Approximate Coordinates	Hairs	Nest Age
14	DEFR10	DOUALA-EDEA/TISSONGO	N 03°35.820' E 09°53.649'	6	5
14	DEFR11	DOUALA-EDEA/TISSONGO	N 03°35.820' E 09°53.649'	15	5
15	CAMPO1	CAMPO RESERVE	N 02°19.377' E 10°08.401'	11	2
15	CAMPO2	CAMPO RESERVE	N 02°19.377' E 10°08.401'	14	2
15	CAMPO3	CAMPO RESERVE	N 02°19.377' E 10°08.401'	16	4
15	CAMPO4	CAMPO RESERVE	N 02°19.377' E 10°08.401'	9	3
15	CAMPO5	CAMPO RESERVE	N 02°19.377' E 10°08.401'	17	4
15	CAMPO6	CAMPO RESERVE	N 02°19.377' E 10°08.401'	15	1
15	CAMPO7	CAMPO RESERVE	N 02°19.377' E 10°08.401'	11	3
15	CAMPO8	CAMPO RESERVE	N 02°19.377' E 10°08.401'	10	3
15	CAMPO9	CAMPO RESERVE	N 02°19.377' E 10°08.401'	45	3
15	CAMP010	CAMPO RESERVE	N 02°19.377' E 10°08.401'	12	3
15	CAMPO11	CAMPO RESERVE	N 02°19.377' E 10°08.401'	7	3
15	CAMPO12	CAMPO RESERVE	N 02°19.377' E 10°08.401'	40	3
15	CAMPO13	CAMPO RESERVE	N 02°19.377' E 10°08.401'	47	2
15	CAMPO14	CAMPO RESERVE	N 02°19.377' E 10°08.401'	24	2
15	CAMPO15	CAMPO RESERVE	N 02°19.377' E 10°08.401'	27	3
16	DBR1	DJA RESERVE	N 03°11.403' E 12°48.706'	8	3
16	DBR2	DJA RESERVE	N 03°11.403' E 12°48.706'	15	3
16	DBR3	DJA RESERVE	N 03°11.403' E 12°48.706'	10	3
16	DBR4	DJA RESERVE	N 03°11.403' E 12°48.706'	2	1
16	DBR5	DJA RESERVE	N 03°11.403' E 12°48.706'	14	2
16	DBR6	DJA RESERVE	N 03°11.403' E 12°48.706'	17	1
16	DBR7	DJA RESERVE	N 03°11.403' E 12°48.706'	15	3
16	DBR8	DJA RESERVE	N 03°11.403' E 12°48.706'	25	1
16	DBR9	DJA RESERVE	N 03°11.403' E 12°48.706'	20	3
16	DBR10	DJA RESERVE	N 03o11.403' E 12o48.706'	25	1
16	DBR11	DJA RESERVE	N 03o11.403' E 12o48.706'	6	4
16	DBR12	DJA RESERVE	N 03o11.403' E 12o48.706'	18	2
16	DBR13	DJA RESERVE	N 03o11.403' E 12o48.706'	6	2
16	DBR14	DJA RESERVE	N 03o11.403' E 12o48.706'	22	4

Map	Sample	Location	Approximate Coordinates	Hairs	Nest Age
16	DBR15	DJA RESERVE	N 03o11.403' E 12o48.706'	17	4
16	DBR16	DJA RESERVE	N 03o11.403' E 12o48.706'	3	1
Skin Samples					
	Sample	Suspected Provenance		Quantity	
5	Abong-SK	ABONG		20g	
2	Ala-SK	ALA		25g	
2	Owo-SK	OWO		15g	
1	Ise-sk	ISE		5g	
Niger Delta	Bomadi-SK	BOMADI--KPAKIEMME		15g	

Captive Samples					
Calabar, Nigeria			Limbe, Cameroon		
Sample	Suspected Provenance	Hairs	Sample	Suspected Provenance	Hairs
Osang	Nigeria	15	Akwaya-Jean	Akwaya	15
Abu	Nigeria	20	Bally	Luna Park, Yaounde	20
Utangah	Nigeria	33	Carlos	Nguti	18
Mickey	Western Nigeria?	10	Damien	Fontem, Tinto	5
Jessica	Nigeria	30	Dja	Dja	4
Daisy	Nigeria	15	Etiange	Muemi Hotel, Yaounde	20
Pansy	Nigeria	10	Ewake	Mamfe	21
Murphy	Nigeria	12	Jack	arrived with Dja from Yaounde	14
Shirley	Nigeria	11	Jacob	Semi New Beach Hotel	12
Po	Nigeria	9	Jules	Bertoua	10
Jaki	Nigeria	14	Louisa	Semi New Beach Hotel	12
Itambu	Nigeria	34	Mac	SEFAC Office, Douala	3
Trog	Nigeria	17	Margaret	Debencher	8
Paablo	Equatorial Guinea	22	Maya	Yaounde Zoo	20
Kingsley	Nigeria	50	Misang	Mt Kupe	4
			Mokolo	conficated from Douala	17
			Poldina	Bertoua	20
			Somba	Yaounde	9
			Suzi	???	4
			Seko	Dja	23
			Tobi	Yaounde Zoo	11

APPENDIX 2

HV-I DNA Sequences used for Analysis

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

Human           AGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGT   [50]
H45             GGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGT   [50]
Ppaniscus       AGTGCCACCCAAGTATTGGCTCATTCACTA-TAACCGCTATGTATTTCGT   [49]
P_1Kp           AGTGCCACCCAAGTATTGGCTCATTCACTA-TAACCGCTATGTATTTCGT   [49]

CRNP2           AAGTACCACCTAAGTACTGGCTCATTCATTACAACCGCTATGTATTTCGT   [50]
CRNP3           ..................................................   [50]
CRNP6           ........??......?..??....?????....................   [50]
CRNP9           ..................................................   [50]
CRNP10          .................................................C   [50]
CRNP12          ................T.....T...................G......G   [50]
AKZN6           ..................................................   [50]
AKZN7           .........?................................AT......   [50]
AKZN8           ...G.......T..G............................G.....G   [50]
AKZN10          ..................................................   [50]
GONCOMP01       ..................................................   [50]
AKZN13          ................T...T.............................   [50]
AKZN14          .......CA.........................................   [50]
AKZN16          ..................................................   [50]
NGNY14          .................................................G   [50]
NGNY15          ...................A..............................   [50]
GGNP951         ............G.......?.............................   [50]
GGNP954         ....................?..........................T..   [50]
GGNP955         ....................?.............................   [50]
GGNP21          .....T......................................A....G   [50]
GGNP23          ..................................................   [50]
GGNP27          .................................................G   [50]
BYM1            ......T...........................................   [50]
BYM3            ................T.....T...........................   [50]
BYM4            ..................................................   [50]
BYM5            ...G................A.............................   [50]
MTCM2           ..................................................   [50]
MTCM4           ..................................................   [50]
MTCM7           ................................?...CG...T.?A....G   [50]
MTCM8           ..................................................   [50]
MTCM11          ..................................................   [50]
MTCM12          ..................................................   [50]
MTCM15          ..................................................   [50]
MTCM16          ..................................................   [50]
MTCM21          ..................................................   [50]
Mosse1          ....?.............................................   [50]
Mosse2          ............................?......A..............   [50]
```

```
Mosse5        ............................?.....................   [50]
Mosse8        ..................................................   [50]
Mosse12       ..................................................   [50]
Mosse13       ..............................G...................   [50]
Mosse14       .................................................G   [50]
Mosse15       .................................................G   [50]
Mosse21       ..................................................   [50]
Man1          ................T.....T...........................   [50]
Man3          ................T.....T...........................   [50]
Man6          ................TG.CT.............................   [50]
Man7          ........G........................................C   [50]
Man8          ................T.....T...........................   [50]
Man10         ...G...............A......................G.......   [50]
Man11         ....................................AG...........G   [50]
Man12         ..................................................   [50]
Man15         ..............................G...................   [50]
Man16         ..........................................G......G   [50]
DEFR1         ................T.....T...........................   [50]
DEFR2         ................T.....T......?.......G............   [50]
DEFR3         .........?......T.....?..?...........G............   [50]
DEFR6         ................T....CT...........................   [50]
DEFR7         ................T.....T...........................   [50]
DEFR8         ................T.....-...........................   [49]
DEFR9         ................T.....T...........................   [50]
Campo1        ...................CT.T...........................   [50]
Campo2        ................T.....T...........................   [50]
Campo6        ......?.........T.....T...........................   [50]
GONCOMP02     ................T.....T...........................   [50]
Campo9        ................T.....T...T.......................   [50]
Campo13       ................T.....T...........................   [50]
Campo14       ................T.....T...........................   [50]
Campo15       ................T...TCT...........................   [50]
DBR2          ................T....CT...........................   [50]
DBR3          ................T.....T...........................   [50]
DBR5          ................T....CT...........................   [50]
GONCOMP03     ................T....CT...........................   [50]
DBR7          ..............ACT-........................G.......   [49]
DBR10         ................T.....T...........................   [50]
DBR12         ................T.....T..........................G   [50]
DBR13         ................T....CT..........................G   [50]
ISEFR1        ...................A..............................   [50]
ISEFR3        ....?..............AG...........................AG   [50]
Gimble        ?...............T.....T...........................   [50]
MORCOMP01     ???.............T.....T...........................   [50]
MORCOMP02     ???.............T.....T....G......................   [50]
MORCOMP03     ?...............T.....T.C.........................   [50]
WL_3          ?????.........??????..T....G......................   [50]
MORCOMP04     ????..........?????...T.....................C..T..   [50]
George        ???????.......?????...T...........................   [50]
Pts_43        ??............??T.....T...........................   [50]
MORCOMP05     ??..............T.....T...........................   [50]
Gemini        ??..............T.....T...........................   [50]
Ivindo        ??..............T.....T...........................   [50]
Makokou       ??..............T.....T......................C....   [50]
```

```
Mopia        ??..............T.....T...........................    [50]
Mouilla      ??..............T.....T...........................    [50]
Ponia        ??..............T.....T...........................    [50]
Ptt_36       ??..............T?...CT...........................    [50]
Ptt_37       ?...............T?....T...........................    [50]
Berthe       ??..............T....CT...........................    [50]
Milla        ??.?..........?????..CT..........?................    [50]
C1           ?...............T.....T...........................    [50]
Ptv_4        ???......................C........................    [50]
Ptv_5        ????????..........................................    [50]
Ptv_15       ???...............................................    [50]
C2           ?.................................................    [50]
C3           ?G................................................    [50]
GAGCOMP01    ?????.............................................    [50]
GAGCOMP02    ?????.............................................    [50]
Ulsse8430    ?????.............................................    [50]
RICCInc58    ?????.............................................    [50]
GAGCOMP03    ?????.............................................    [50]
13_p84339    ?????...............................?.............    [50]
12_84303     ??????????????..?...??.??......?..................    [50]
GAGCOMP04    ?????.............................................    [50]
GAGCOMP05    ?????.............................................    [50]
Ali          ?????.....??....?........C........................    [50]
GAGCOMP6     ???????...........................................    [50]
Darwin       ?????.............................................    [50]
Dilly        ?????..............................?.?............    [50]
Hera         ???????........?..??..............................    [50]
Loukoum      ??????..........?.................................    [50]
Ondine       ?????..............................?.?............    [50]
Perla        ??????..?..?....?.................................    [50]
Rousseau     ?????.............................................    [50]
Salome       ??????????????....................................    [50]
Venus        ?????????.........................................    [50]
Xeres        ?????.............................................    [50]
B3           ?????????.........................................    [50]
F18          ?????...........?.................................    [50]
K22          ???????.........?.................................    [50]
M19          ??????..?.......?........??????...................    [50]
M18          ??????????????....................................    [50]
Solo10       ?????..............................?.?............    [50]
Solo13       ?????......??........??...........................    [50]
Wongo1       ?????????......????...............................    [50]
Wongo7       ??????...?...?.?...?...?..........................    [50]
Wongo5       ??????...?.....?...?..............................    [50]
Pts001.      ................T.....T...........................    [50]
Pts002       ................T.....T...........................    [50]
Pts003       ................T.....T...........................    [50]
Pts005.      ???.............T.....T..........?................    [50]
Pts006.      ................T.....T...........................    [50]
Pts007.      ................T.....T...........................    [50]
Pts008.      ................T.....T...........................    [50]
Pts009.      ................T.....T...........................    [50]
Pts010.      ................T.....T...........................    [50]
Pts011.      ................T.....T...........................    [50]
```

```
Pts012.     ................T.....T...........................   [50]
Pts013.     ................T.....T...........................   [50]
Pts014.     ................T.....T...........................   [50]
Pts015.     ................T.....T...........................   [50]
Pts016.     ................T.....T...........................   [50]
Pts017.     ................T.....T.....C.....................   [50]
Pts018.     ???.........?...T.....T...........................   [50]
Pts019.     ................T.....T...........................   [50]
Pts020.     ................T.....T...........................   [50]
Pts021.     ................T.....T...........................   [50]
Pts022.     ................T.....T...........................   [50]
Pts023.     ................T.....T...........................   [50]
Pts024.     ................T.....T...........................   [50]
Pts025.     ?...............T.....T.C.........................   [50]
Pts026.     ................T.....T...........................   [50]
Pts027.     ................T.....T...........................   [50]
Pts028.     ................T.....T...........................   [50]
Pts029.     ................T.....T....G......................   [50]
Pts030.     ................T.....T....G......................   [50]
Pts031.     ?????.........??????..T....G......................   [50]
Pts032.     ???.............T.....T....G......................   [50]
Pts033.     ................T.....T....G......................   [50]
Pts034.     ................T.....T....G......................   [50]
Pts035.     ................T.....T.....................C.....   [50]
Pts036.     ................T.....T...........................   [50]
Pts037.     ................T.....T...........................   [50]
Pts038.     ................T.....T...........................   [50]
Pts039.     ................T.....T...........................   [50]
Pts040.     ................T.....T...........................   [50]
Pts041.     ????..........?????...T.....................C..T..   [50]
Pts043.     ................T.....T...........................   [50]
Pts044.     ................T.....T...........................   [50]
Pts045.     ................T.....T...........................   [50]
Pts046.     ................T.....T.............A.............   [50]
Pts047.     ................T.....T...........................   [50]
Pts048.     ................T.....T......................C....   [50]
Pts049.     ...........G....T.....T..........G..........C.....   [50]
Pts050.     ................T.....T...........................   [50]
Pts051.     ................T.....T...........................   [50]
Pts052.     ................T.....T...........................   [50]
Pts053.     ................T.....T...........................   [50]
Pts054.     ................T.....T...........................   [50]
Pts055.     ................T.....T...........................   [50]
Pts056.     ................T.....T...........................   [50]
Pts057.     ................T.....T...........................   [50]
Pts058.     ................T.....T...........................   [50]
Pts059.     ................T.....T...........................   [50]
Pts060.     ................T.....T...........................   [50]
Pts061.     ................T.....T...........................   [50]
Pts062.     ................T.....T...........................   [50]
Pts063.     ................T.....T...........................   [50]
Pts065.     ................T.....T...........................   [50]
Pts066.     ................T.....T...........................   [50]
Pts067.     ................T.....T...........................   [50]
```

```
Pts068.      ................T.....T...........................   [50]
Pts069.      ................T.....T...........................   [50]
Pts070.      ................T.....T...........................   [50]
Pts071.      ................T.....T...........................   [50]
Pts072.      ................T.....T...........................   [50]
Pts073.      ................T.....T...........................   [50]
Pts074.      ................T.....T...........................   [50]
Pts075.      ................T.....T...........................   [50]
Pts076.      ................T.....T...........................   [50]
Pts077.      ................T.....T...........................   [50]
Pts078.      ................T.....T...........................   [50]
Pts079.      ................T.....T...........................   [50]
Pts080.      ..............C.T.....T...........................   [50]
Pts081.      ..............C.T.....T...........................   [50]
Pts082.      ..............C.T.....T...........................   [50]
Pts083.      ................T.....T...........................   [50]
Pts084.      ................T.....T...........................   [50]
Pts085.      ................T.....T...........................   [50]
Pts086.      ................T.....T...........................   [50]
GOLCOMP01    ................T.....T...........................   [50]
Pts089.      ................T.....T...........................   [50]
Pts090.      ................T.....T...........................   [50]
Pts091.      ................T.....T...........................   [50]
Pts092.      ................T.....T...........................   [50]
Pts093.      ................T.....T...........................   [50]
Pts094.      ................T.....T...........................   [50]
Pts095.      ................T.....T...........................   [50]
Pts096.      ................T.....T...........................   [50]
Pts097.      ................T.....T...........................   [50]
Pts098.      ................T.....T...........................   [50]
Pts099.      ................T.....T...........................   [50]
Pts100.      .....T..........T.....T...........................   [50]
Pts101.      ................T.....T...........................   [50]
Pts102.      ................T.....T...........................   [50]
Pts103.      ................T.....T...........................   [50]
Pts104.      ................T.....T...........................   [50]
Pts105.      ................T.....T...........................   [50]
Pts106.      ................T.....T...........................   [50]
Pts107.      ................T.....T.....................C.....   [50]
Pts108.      ................T.....T...........................   [50]
Pts109.      ................T.....T...........................   [50]
Pts110.      ................T.....T...........................   [50]
Pts111.      ................T.....T......................C....   [50]
Pts112.      ................T.....T...........................   [50]
Pts113.      ................T.....T...........................   [50]
Pts114.      ................T.....T...........................   [50]
Pts115.      ................T.....T...........................   [50]
Pts116.      ................T.....T...........................   [50]
Pts117.      ................T.....T...........................   [50]
Pts118.      ................T.....T...........................   [50]
Pts119.      ................T.....T...........................   [50]
Pts120.      ................T.....T...........................   [50]
Pts121.      ?...............T.....T.......?...................   [50]
Pts122.      ................T.....T...........................   [50]
```

```
Pts123.         ................T.....T...........................   [50]

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

Human           ACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGAC   [100]
H45             ACATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGAC   [100]
Ppaniscus       ACATTACTGCCAGCCACCATGAATATTACATAGTACTATAATCATTTAAC   [99]
P_1Kp           ACATTACTGCCAGCCACCATGAATATTACATAGTACTATAATCATTTAAC   [99]

CRNP2           ACATTACTGCCAGTCACCATGAATATTGTACAGTACCACAATCACTCAAC   [100]
CRNP3           .............C........................T...........   [100]
CRNP6           .............C......................T.T...........   [100]
CRNP9           .............C....................................   [100]
CRNP10          ..........A..C........................T...........   [100]
CRNP12          .............C......................T.T...........   [100]
AKZN6           .............C....................................   [100]
AKZN7           ..................................................   [100]
AKZN8           ..................................................   [100]
AKZN10          .............C...............................CT...   [100]
GONCOMP01       .............C...............................CT...   [100]
AKZN13          .............C...................G................   [100]
AKZN14          ..................................................   [100]
AKZN16          ..................................................   [100]
NGNY14          ..................................................   [100]
NGNY15          ..................................................   [100]
GGNP951         ..................................................   [100]
GGNP954         .............C.......?.......................CT...   [100]
GGNP955         ..................................................   [100]
GGNP21          ..................................................   [100]
GGNP23          .............C........................T......C....   [100]
GGNP27          .............C...............................CT...   [100]
BYM1            ..................................................   [100]
BYM3            ..........T..C............C.........T.T...........   [100]
BYM4            .............C....................................   [100]
BYM5            .............C....................................   [100]
MTCM2           .............C............C..................C....   [100]
MTCM4           .........G...C........................T......C....   [100]
MTCM7           ....?........C........................T......C....   [100]
MTCM8           ..................................................   [100]
MTCM11          .............C........................T......C....   [100]
MTCM12          ..................................................   [100]
MTCM15          .............C............C..................C....   [100]
MTCM16          .............C....................................   [100]
MTCM21          .............C........................T......C....   [100]
Mosse1          .............C....................................   [100]
Mosse2          .............C....................................   [100]
Mosse5          ..................................................   [100]
Mosse8          ..................................................   [100]
Mosse12         .............C........................T......C....   [100]
Mosse13         .............C....................................   [100]
Mosse14         .............C........................T......C....   [100]
Mosse15         ..................................................   [100]
Mosse21         .............C....................................   [100]
```

```
Man1        .............C......................T.T.G.........    [100]
Man3        .............C......................T.T.G.........    [100]
Man6        .............C........................T......C....    [100]
Man7        .............CT...................................    [100]
Man8        .............C......................T.T...........    [100]
Man10       .............C....................................    [100]
Man11       .............C....................................    [100]
Man12       .............................C....................    [100]
Man15       .............C....................................    [100]
Man16       T............C....................................    [100]
DEFR1       .............C......................T.T...........    [100]
DEFR2       .............C.C....................T.T...........    [100]
DEFR3       .............C.C....................T.T...........    [100]
DEFR6       .............C......................T.T..C........    [100]
DEFR7       .............C......................T.T...........    [100]
DEFR8       .............C......................T.T...........    [99]
DEFR9       .............C......................T.T...........    [100]
Campo1      .............C......................T.T...........    [100]
Campo2      .............C......................T.T......C....    [100]
Campo6      .............C......................T.T......C....    [100]
GONCOMP02   .............C......................T.T...........    [100]
Campo9      .............C......................T.T...A.......    [100]
Campo13     .............C......................T.T...........    [100]
Campo14     .............C......................T.T...........    [100]
Campo15     .............C......................T.T...........    [100]
DBR2        .............C......................T.T...........    [100]
DBR3        .........G...C......................T.T...........    [100]
DBR5        .............C...................G..T.T...........    [100]
GONCOMP03   .............C......................T.T...........    [100]
DBR7        .............C............C..................C....    [99]
DBR10       .............C......................T.T...........    [100]
DBR12       .............C......................T.T...........    [100]
DBR13       .............C......................T.T..C........    [100]
ISEFR1      .............CT...T..............................T    [100]
ISEFR3      .............C...................................T    [100]
Gimble      .............C......................T.T...........    [100]
MORCOMP01   .............C......................T.T.....T.....    [100]
MORCOMP02   .............C......................T.T...........    [100]
MORCOMP03   .............C......................T.T...........    [100]
WL_3        .............C...................?..T.T...........    [100]
MORCOMP04   .............C......................T.T...........    [100]
George      .............C......................T.T...........    [100]
Pts_43      .............C......................T.T...........    [100]
MORCOMP05   .............C......................T.T......C....    [100]
Gemini      .............C......................T.T...........    [100]
Ivindo      .............C......................T.T...........    [100]
Makokou     .............C......................T.T...........    [100]
Mopia       .............C......................T.T...........    [100]
Mouilla     .............C......................T.T...........    [100]
Ponia       .............C......................T.T...........    [100]
Ptt_36      .............C......................T.T...........    [100]
Ptt_37      .............C......................T.T......C....    [100]
Berthe      .............C......................T.T...........    [100]
Milla       .............C......................T.T......C....    [100]
```

```
C1           .............C......................T.T...........   [100]
Ptv_4        .............C.................G......T...T..C....   [100]
Ptv_5        .............C........................T......C....   [100]
Ptv_15       .............C........................T......C....   [100]
C2           .............C.................G......T...T..C....   [100]
C3           .............C............C...........T.-....C....   [99]
GAGCOMP01    .............C.................G......T...T..C....   [100]
GAGCOMP02    .............C........................T......C....   [100]
Ulsse8430    .............C........................T......C....   [100]
RICCInc58    .............C............?...........T......C....   [100]
GAGCOMP03    .............C............C...........T......C....   [100]
13_p84339    .............C........................T......C....   [100]
12_84303     .............C........................T......C....   [100]
GAGCOMP04    ......................................T......C....   [100]
GAGCOMP05    .............C.................G......T......C....   [100]
Ali          .............C........................T...T..C....   [100]
GAGCOMP6     .............C........................T......C....   [100]
Darwin       .........T............................T......C....   [100]
Dilly        .............C.................G......T...T..C....   [100]
Hera         .............C............?...........T......C....   [100]
Loukoum      .............C.................G......T......C....   [100]
Ondine       .............C........................T......C....   [100]
Perla        .............C.................G......T...T..C....   [100]
Rousseau     ......................................T......C....   [100]
Salome       .............C............C...........T......C....   [100]
Venus        .............C........................T......C....   [100]
Xeres        .............C........................T......C....   [100]
B3           .............C.................G......T......C....   [100]
F18          .............C.................G......T...T..C....   [100]
K22          .............C.................G......T......C....   [100]
M19          .............C.................G......T...T..C....   [100]
M18          .............C.................G......T......CT...   [100]
Solo10       .............C.................G......T......C....   [100]
Solo13       .........T...C........................T......C.?..   [100]
Wongo1       .............C........................T......C....   [100]
Wongo7       .............C........................T......C....   [100]
Wongo5       .............C........................T......C....   [100]
Pts001.      .............C......................T.T...........   [100]
Pts002       .............C......................T.T...........   [100]
Pts003       .............C......................T.T...........   [100]
Pts005.      .............C........?.............T.T.....T.....   [100]
Pts006.      .............C......................T.T.....T.....   [100]
Pts007.      .............C......................T.T.....T.....   [100]
Pts008.      .............C......................T.T.....T.....   [100]
Pts009.      .............C......................T.T.....T.....   [100]
Pts010.      .............C......................T.T.....T.....   [100]
Pts011.      .............C......................T.T...........   [100]
Pts012.      .............C......................T.T...........   [100]
Pts013.      .............C......................T.T...........   [100]
Pts014.      .............C......................T.T...........   [100]
Pts015.      .............C......................T.T...........   [100]
Pts016.      .............C......................T.T...........   [100]
Pts017.      .............C......................T.T.....T.....   [100]
Pts018.      .............C......................T.T.....T.....   [100]
```

```
Pts019.    .............C......................T.T...........   [100]
Pts020.    .............C......................T.T...........   [100]
Pts021.    .............C......................T.T...........   [100]
Pts022.    .............C......................T.T...T.......   [100]
Pts023.    .............C........................T...T.......   [100]
Pts024.    .............C......................T.T...T.......   [100]
Pts025.    .............C......................T.T...........   [100]
Pts026.    .............C......................T.T...T.......   [100]
Pts027.    .............C......................T.T...........   [100]
Pts028.    ....................................T.T...........   [100]
Pts029.    .............C......................T.T...........   [100]
Pts030.    .............C......................T.T...........   [100]
Pts031.    .............C..........?........??.T.T...........   [100]
Pts032.    .............C......................T.T...........   [100]
Pts033.    .............C......................T.T...........   [100]
Pts034.    .............C......................T.T...........   [100]
Pts035.    .............C......................T.T...........   [100]
Pts036.    .............C......................T.T...........   [100]
Pts037.    .............C......................T.T...........   [100]
Pts038.    .............C......................T.T...........   [100]
Pts039.    .............C......................T.T...........   [100]
Pts040.    .............C......................T.T...........   [100]
Pts041.    .............C......................T.T...........   [100]
Pts043.    .............C......................T.T...........   [100]
Pts044.    .............C......................T.T...........   [100]
Pts045.    .............C......................T.T...........   [100]
Pts046.    ....................................T.T...........   [100]
Pts047.    .............C......................T.T...........   [100]
Pts048.    .............C......................T.T...........   [100]
Pts049.    .............C......................T.............   [100]
Pts050.    ....................................T.T...........   [100]
Pts051.    ....................................T.T...........   [100]
Pts052.    .............C......................T.T...........   [100]
Pts053.    .............C......................T.T...........   [100]
Pts054.    .............C......................T.T...........   [100]
Pts055.    .............C......................T.T...........   [100]
Pts056.    .............C......................T.T...........   [100]
Pts057.    .............C......................T.T...........   [100]
Pts058.    .............C......................T.T...........   [100]
Pts059.    ....................................T.T...........   [100]
Pts060.    ....................................T.T...........   [100]
Pts061.    .............C......................T.T...........   [100]
Pts062.    .............C......................T.T...........   [100]
Pts063.    .............C......................T.T...........   [100]
Pts065.    .............C......................T.T...........   [100]
Pts066.    .............C............C.........T.T...........   [100]
Pts067.    .............C......................T.T...........   [100]
Pts068.    .............C......................T.T...........   [100]
Pts069.    .............C......................T.T...........   [100]
Pts070.    .............C......................T.............   [100]
Pts071.    .............C......................T.T...........   [100]
Pts072.    .............C......................T.T...........   [100]
Pts073.    .............C.............A........T.T...........   [100]
Pts074.    .............C......................T.T...........   [100]
```

```
Pts075.         .............C......................T.T...........   [100]
Pts076.         .............C......................T.T...........   [100]
Pts077.         .............C......................T.T...........   [100]
Pts078.         .............C......................T.T...........   [100]
Pts079.         .............C......................T.T...........   [100]
Pts080.         .............C......................T.T...........   [100]
Pts081.         .............C......................T.T...........   [100]
Pts082.         .............C......................T.T...........   [100]
Pts083.         .............C......................T.T...........   [100]
Pts084.         .............C......................T.T...........   [100]
Pts085.         .............C......................T.T...........   [100]
Pts086.         .............C......................T.T...........   [100]
GOLCOMP01       .............C......................T.T...........   [100]
Pts089.         .............C......................T.T...........   [100]
Pts090.         .............C......................T.T...........   [100]
Pts091.         .............C......................T.T...........   [100]
Pts092.         .............C......................T.T...........   [100]
Pts093.         ....................................T.T...........   [100]
Pts094.         .........T...C......................T.T...........   [100]
Pts095.         .............C......................T.T...........   [100]
Pts096.         .............C......................T.T...........   [100]
Pts097.         .............C......................T.T...........   [100]
Pts098.         .............C......................T.T...........   [100]
Pts099.         .............C......................T.T...........   [100]
Pts100.         .............C......................T.T...........   [100]
Pts101.         .............C......................T.T...........   [100]
Pts102.         .............C......................T.T...........   [100]
Pts103.         .............C......................T.T...........   [100]
Pts104.         .............C......................T.T...........   [100]
Pts105.         .............C......................T.............   [100]
Pts106.         .............C......................T.............   [100]
Pts107.         .............C......................T.............   [100]
Pts108.         .............C......................T.............   [100]
Pts109.         .............C......................T.............   [100]
Pts110.         .............C......................T.T...........   [100]
Pts111.         .............C......................T.T...........   [100]
Pts112.         .............C......................T.T...........   [100]
Pts113.         .............C......................T.T...........   [100]
Pts114.         .............C......................T.T...........   [100]
Pts115.         .............C......................T.T...........   [100]
Pts116.         .............C......................T.T...........   [100]
Pts117.         .............C......................T.T...........   [100]
Pts118.         .............C......................T.T...........   [100]
Pts119.         .............C......................T.T...........   [100]
Pts120.         .............C......................T.T...........   [100]
Pts121.         ?............C......................T.T.?.........   [100]
Pts122.         .............C......................T.T...........   [100]
Pts123.         .............C......................T.T...........   [100]

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

Human           CACCTGTAGTACATAAAAACCCAATCC-ACATCAAAACCCCCCCCCCATG   [149]
H45             CACCTGTAGTACATAAAAACCCAATCC-ACATCAAAACCCTCACCCCATG   [149]
```

```
Ppaniscus    CACCTATAACACATAAAAACCTACATCCACATTAAAACCCCCCCCCCATG   [149]
P_1Kp        CACCTATAACACATAAAAACCTACATCCACATTAAAACCCCCCCCCCATG   [149]

CRNP2        TACCTATAACACATAAAACCCACCCCC-ACATCAAAATCTCCACCCCATG   [149]
CRNP3        .........T......C..........-........GC..T.........   [149]
CRNP6        ..............C........T..--....T.C..CG...T.......   [148]
CRNP9        ...........................-.........CT...........   [149]
CRNP10       .........T......C..........-........GC..T.........   [149]
CRNP12       ..............C........T..--....TGC..CG...C.......   [148]
AKZN6        ...........................-.........C............   [149]
AKZN7        ...........................-.........C............   [149]
AKZN8        ...........................-.........C............   [149]
AKZN10       .........T..........T....T.-.........G..........C.   [149]
GONCOMP01    .........T..........T....T.-.........G..........C.   [149]
AKZN13       ...........................-.........C............   [149]
AKZN14       .........T.................-.........CT...........   [149]
AKZN16       ...........................-.........C............   [149]
NGNY14       .........T.................-.........C............   [149]
NGNY15       .........T.................-.........C............   [149]
GGNP951      .........T.................-.........CT...........   [149]
GGNP954      .........T..........T....T.-.........G..........C.   [149]
GGNP955      .........T.................-.........CT...........   [149]
GGNP21       ...........................-.........C............   [149]
GGNP23       ...........................-.........C.C.....T....   [149]
GGNP27       .........T..........T....T.-.........G..........C.   [149]
BYM1         ...........................-.........C............   [149]
BYM3         C........T....C...........--....T.C..C..T.........   [148]
BYM4         ...........................-.........C............   [149]
BYM5         ...........................-.........C............   [149]
MTCM2        ......C..T..........T......-.........G............   [149]
MTCM4        ..................T........-.........C............   [149]
MTCM7        ..................T........-.........C............   [149]
MTCM8        ...........................-......................   [149]
MTCM11       ..................T........-.........C............   [149]
MTCM12       ...........................-.........C............   [149]
MTCM15       ......C..T..........T......-.........G............   [149]
MTCM16       ...........................-.........C............   [149]
MTCM21       ...........................-.........C.......T....   [149]
Mosse1       .........T.................-.........C............   [149]
Mosse2       .........T.................-.........C............   [149]
Mosse5       ...........................-.........C............   [149]
Mosse8       ...........................-.........C............   [149]
Mosse12      ...........................-.........C............   [149]
Mosse13      .........T.................-.........C............   [149]
Mosse14      ...........................-.........C............   [149]
Mosse15      ...........................-.........C............   [149]
Mosse21      .........T.................-.........C............   [149]
Man1         C.............T........TT.--....T.C..CG...C.......   [148]
Man3         C.............T........TT.--....T.C..CG...C.......   [148]
Man6         ...........................-.........C.......T....   [149]
Man7         ...........................-.........C............   [149]
Man8         C.............T........TT.--....T.C..CG...C.......   [148]
Man10        ...........................-.........C............   [149]
Man11        ...........................-.........C............   [149]
```

```
Man12       ...........................-.........C............   [149]
Man15       .........T.................-.........C............   [149]
Man16       ...........................-.........C............   [149]
DEFR1       ..............C........T..--....T.C..CG...T.......   [148]
DEFR2       ..............C........T..--....T.C..CG...T.......   [148]
DEFR3       ..............C........T..--....TGC..CG...C.......   [148]
DEFR6       .........T....T............-....T.C.......C...T...   [149]
DEFR7       ..............T........T..--....T.C..CA...C.......   [148]
DEFR8       ..............T........T..--....T.C..CG...C.......   [147]
DEFR9       ..............C........T..--....T.C..CG...T.......   [148]
Campo1      ..............C........T..--....TGC..CG...C.......   [148]
Campo2      ..............T...........--....T.C..CG...C.......   [148]
Campo6      ..............T...........--....T.C..CG...C.......   [148]
GONCOMP02   ..............C........T..--....TGC..CG...C.......   [148]
Campo9      .........T....C...........--....T.C..C............   [148]
Campo13     ..............C........T..--....T.C..CG...C.......   [148]
Campo14     ..............C........T..--....TGC..CG...C.......   [148]
Campo15     ..............C........T..--....T.C..CG...C.......   [148]
DBR2        .........T....T............-....A.C..C........T...   [149]
DBR3        .........T....T............-....A.C..C........T...   [149]
DBR5        .........T....T............-....A.C..C........T...   [149]
GONCOMP03   .........T....T............-....A.C..C........T...   [149]
DBR7        ......C..T..........T......-.........G............   [148]
DBR10       ..............C.........T.--....T.C...G...C.......   [148]
DBR12       ..........G...C........T..--....T.C..CA...C.......   [148]
DBR13       .........T....T............-....T.C.......T...T...   [149]
ISEFR1      .....G............T........-....T....C.C..C.......   [149]
ISEFR3      .....G............T........-....T....C.C..C.......   [149]
Gimble      .........T....C...........--....T.T..C........T...   [148]
MORCOMP01   .........T....C...........--....T.C..C............   [148]
MORCOMP02   .........T....C...........--....T.T..C............   [148]
MORCOMP03   .........T....C...........--....T.C.GC............   [148]
WL_3        .........T....C...........--....T.T..C............   [148]
MORCOMP04   .........T....C....T......--....T.C..C............   [148]
George      .........T....C...........--....T.C..C............   [148]
Pts_43      .........T....C...........--....T.T..C.??T........   [148]
MORCOMP05   C........T....T...........--....T.C..CG...........   [148]
Gemini      ........G.....T........T..--....T.C..CG...C.......   [148]
Ivindo      ..............T...T....T..--....T.C...G...C....CC.   [148]
Makokou     ..............T...........--....T.C..CG...C..T....   [148]
Mopia       ..............T...........--....T.C..CG...C.......   [148]
Mouilla     ..............T........T..--....T.C..CG...C.......   [148]
Ponia       .........T....T........T..--....T.C..CG...C.......   [148]
Ptt_36      .........T....T............C....T.C..C........T...   [150]
Ptt_37      ..............T...........--....T.C..CG...C.......   [148]
Berthe      .........T....T............-....A.C..C........T...   [149]
Milla       .........T....T...........TT....T.C..C........T...   [150]
C1          ..............C.......AT...-....T.C..CG...C.......   [149]
Ptv_4       ..........................--....T...GC..T.........   [148]
Ptv_5       ..................T....T...-...C.....C..T.........   [149]
Ptv_15      .........T........T....T...-...C.....C..T.........   [149]
C2          ......................A....-....T....C..T.........   [149]
C3          ..................T....T...-....T....C..T...T.....   [148]
GAGCOMP01   ..........................--....T....C..T.........   [148]
```

```
GAGCOMP02      ..................T....T...-...C.....C..T.........   [149]
Ulsse8430      .........T.?......T....T...-...C.....C..T.........   [149]
RICCInc58      C.......G.........T.....T.--....T....C..T.........   [148]
GAGCOMP03      ........GT........T....T...-.......G.C..T.........   [149]
13_p84339      .........T........T....T...-...C.....C..T.........   [149]
12_84303       C.......G.........T.....T.--....T....C..T?........   [148]
GAGCOMP04      .........T........T....T...-...C.....C..T.........   [149]
GAGCOMP05      ........G.................--....T....C..T.........   [148]
Ali            ..................T?..?T?..-...C.....??.T??.......   [149]
GAGCOMP6       .........T........T....T...-...C.....C..T.........   [149]
Darwin         .........T........T....?...-...C.....C..T??.......   [149]
Dilly          .................??...?...--....T....C..T.........   [148]
Hera           ........GT........T....T...-....T....C..T?..T??...   [149]
Loukoum        ........G.................--....T....C..T.........   [148]
Ondine         ........G.........T.....T.--....T....C..T.........   [148]
Perla          ..........................--....T..G.C..T.........   [148]
Rousseau       ..................T....T...-...C.....C..T.........   [149]
Salome         ........GT........T....T...-.......G.C..T...T.....   [149]
Venus          ........GT........T....T...-.......G.C..T...T.....   [149]
Xeres          .........T........T....T...-...C.....C..T.???.....   [149]
B3             ........G.................--....T....C..T.........   [148]
F18            ..........................--....T....C..T.........   [148]
K22            ........G.................--....T....C..T.........   [148]
M19            ..........................--....T....C..T.........   [148]
M18            ........G.................--....T....C..T.........   [148]
Solo10         .?.......T........????????.-....T...?C..T.........   [149]
Solo13         ..................T....T...-...C.....C..T.........   [149]
Wongo1         ..........?...?...T....T...-...C??...C?.T.?.......   [149]
Wongo7         ..................T....T...-...C.....C..T.........   [149]
Wongo5         C.......GT........T.....T.--.........C..T.........   [148]
Pts001.        ..............C...T.......--....T.T..C............   [148]
Pts002         ..............C...T.......--....T.T..C............   [148]
Pts003         ..............C...T.......--....T.T..C............   [148]
Pts005.        .........T....C...........--......C..C............   [148]
Pts006.        .........T....C....T......--......C..C............   [148]
Pts007.        .........T....C....T......--......C..C............   [148]
Pts008.        .........T....C....T......--......C..C............   [148]
Pts009.        .........T....C....T......--......C..C............   [148]
Pts010.        .........T....C....T......--......C..C............   [148]
Pts011.        .........T....C...T.......--G...T.T..C...T........   [148]
Pts012.        .........T....C...T.......--G...T.T..C............   [148]
Pts013.        .........T....C...T.......--G...T.T..C............   [148]
Pts014.        .........T....C...........--G...T.T..C............   [148]
Pts015.        .........T....C...........--G...T.C.GC............   [148]
Pts016.        .........T....C...........--G...T.C..C............   [148]
Pts017.        C........T....C....T......--....T.C..C............   [148]
Pts018.        .........T....C...........--....T.C..C............   [148]
Pts019.        .........T....C...........--....T.C.GC............   [148]
Pts020.        .........T....C...........--....T.C.GC............   [148]
Pts021.        .........T....C...........--....T.C.GC............   [148]
Pts022.        ......C..T....C....T......--....T.C.GC............   [148]
Pts023.        .........T....C...........--....T.C.GC............   [148]
Pts024.        .........T....C...........--....T.C.GC............   [148]
Pts025.        .........T....C...........--....T.C.GC............   [148]
```

```
Pts026.     .........T....C...........--....T.C.GC.C..........   [148]
Pts027.     .........T....C...........--....T.C.GC.C..........   [148]
Pts028.     .........T....C.....T.....--....T.T..C...T........   [148]
Pts029.     .........T....C.....T.....--....T.T..C............   [148]
Pts030.     .........T....C.....T.....--....T.T..C............   [148]
Pts031.     .........T....C...........--....T.T.?C............   [148]
Pts032.     .........T....C...........--....T.T..C............   [148]
Pts033.     .........T....C...........--....T.T..C............   [148]
Pts034.     .........T....C...........--....T.T..C............   [148]
Pts035.     .........T....C...........--....T.T..C............   [148]
Pts036.     .........T....C...........--....T.T..C........T...   [148]
Pts037.     .........T....C...........--....T.T..C........T...   [148]
Pts038.     .........T....C...........--....T.T..C............   [148]
Pts039.     .........T....C...........--....T.T..C............   [148]
Pts040.     .........T....C...........--....T.T..C............   [148]
Pts041.     .........T....C....T......--....T.C?.C............   [148]
Pts043.     .........T....C....T......--....T.C..C............   [148]
Pts044.     .........T....C....T......--....T.C..C............   [148]
Pts045.     .........T....T...........--G...T.C..C....C.......   [148]
Pts046.     .........T....C...........--....T.C..C...T........   [148]
Pts047.     .........T....C...........--....T.C..C...T........   [148]
Pts048.     .........T....C...T.......--....T.C..C........T...   [148]
Pts049.     .........T....C...........--....T.T..C............   [148]
Pts050.     .........T....C...........--....T.T..C...T....T...   [148]
Pts051.     .........T....C...........--....T.T..C...T....T...   [148]
Pts052.     .........T....C...........--....T.T..C...T........   [148]
Pts053.     .........T....C...........--....T.T..C...T........   [148]
Pts054.     .........T....C...........--....T.T..C...T........   [148]
Pts055.     .........T....C...........--......T..C...T........   [148]
Pts056.     .........T....C...........--....T.T..C...T....T...   [148]
Pts057.     .........T....C...........--....T.T.GC...T........   [148]
Pts058.     .........T....C...........--....T.T..C...T........   [148]
Pts059.     .........T....C...........--....T.T..C...T........   [148]
Pts060.     .........T....C...........--....T.T..C...T........   [148]
Pts061.     .........T....C....T......--....T.T..C...T........   [148]
Pts062.     .........T....C...........--....T.T..C...T........   [148]
Pts063.     .........T....C...........--....T.T..C...T........   [148]
Pts065.     .........T....C...........--....T.T..C...T........   [148]
Pts066.     .........T....C.....T.....--....T.C.GC........T...   [148]
Pts067.     .........T....C.......T...--....T.T..C...T....T...   [148]
Pts068.     ..............C...........--....T.C..C............   [148]
Pts069.     .........T....C...........--....T.C..C............   [148]
Pts070.     .........T....C...........--....T.C..C...T........   [148]
Pts071.     .........T....C...........--....T.C..C...T........   [148]
Pts072.     .........T....C...........--....T.C..C...T........   [148]
Pts073.     .........T....C...T.......--....T.C..C...T........   [148]
Pts074.     .........T....C...........--....T.C..C...T........   [148]
Pts075.     .........T....C...........--....T.C..C...T........   [148]
Pts076.     .........T....C...........--....T.C..C............   [148]
Pts077.     .........T....C...........--....T.C..C...T........   [148]
Pts078.     .........T....C...........--....T.C..C............   [148]
Pts079.     .........T....C...........--....T.C..C...T........   [148]
Pts080.     .........T....C...........--....T.C..C...T........   [148]
Pts081.     .........T....C...........--....T.C..C...T........   [148]
```

```
Pts082.         .........T....C...........--....T.C..C...T........   [148]
Pts083.         .........T....C...........--....T.C..C...T........   [148]
Pts084.         .........T....C...........--....T.C..C...T........   [148]
Pts085.         .........T....C....T......--....T.T..C......T.....   [148]
Pts086.         .........T....C...........--....T.C..C............   [148]
GOLCOMP01       .........T....C...........--....T.C..C........T...   [148]
Pts089.         .........T....C...........--....T.C..C........T...   [148]
Pts090.         .........T....C...........--....T.C..C...T....T...   [148]
Pts091.         .........T....C...........--....T.C..C........T...   [148]
Pts092.         .........T....C...........--....T.C..C........T...   [148]
Pts093.         .........T....C...........--....T.C..C........T...   [148]
Pts094.         .........T....C...........--....T.C..C........T...   [148]
Pts095.         .........T....C...........--....T.C..C........T...   [148]
Pts096.         .........T....C...........--....T.C..C........T...   [148]
Pts097.         .........T....C...........--....T.T..C........T...   [148]
Pts098.         .........T....C...........--....T.C..C........T...   [148]
Pts099.         .........T....C...........--....T.C..C........T...   [148]
Pts100.         .........T....C...........--....T.C..C........T...   [148]
Pts101.         .........T....C...........--....T.C..C........T...   [148]
Pts102.         .........T....C...........--....T.C..C........T...   [148]
Pts103.         .........T....C...T....T..--....T.T..C........T...   [148]
Pts104.         .........T....C...........--....T.T..C...T....T...   [148]
Pts105.         .........T....C...........--....T.T..C............   [148]
Pts106.         .........T....C...........--....T.T..C............   [148]
Pts107.         .........T....C...........--....T.T..C............   [148]
Pts108.         .........T....C...........--....T.T..C............   [148]
Pts109.         .........T....C...........--....T.T..C............   [148]
Pts110.         .........T....C...........--....T.T..C........T...   [148]
Pts111.         .........T....C...........--....T.T..C............   [148]
Pts112.         .........T....C...........--....T.T..C............   [148]
Pts113.         .........T....C...........--......T..C............   [148]
Pts114.         .........T....C...........--....T.T..C............   [148]
Pts115.         .........T....C...........--....T.T..C............   [148]
Pts116.         .........T....C...........--....T.T..C........T...   [148]
Pts117.         .........T....C...........--....T.C..C............   [148]
Pts118.         .........T....C...........--....T.C..C............   [148]
Pts119.         .........T....C...........--....T.C..C........T...   [148]
Pts120.         .........T....C...........--....T.C..C........T...   [148]
Pts121.         .........T?...C...........--....T.T..C....?...T...   [148]
Pts122.         .........T....C...........--....T.T..C............   [148]
Pts123.         .........T....C...........--....T.T..C............   [148]

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

Human           CTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCA   [199]
H45             CTTACAAGCAAGTACAGTAATCAACCCTCAACTATCATACATCAACTGCA   [199]
Ppaniscus       CATATAAGCACGAACAATAATCAACCTCCAACTGTCGAACAT-AAACACC   [198]
P_1Kp           CATATAAGCACGAACAATAATCGACCTCCAACTGTCGAACAT-AAACACC   [198]

CRNP2           CTTACAAGCACGCACAATAATCAACCCCCAACTGTCGAACATAAAACACA   [199]
CRNP3           .................C.....C..........................   [199]
CRNP6           ............T....C.........T........AC.......G.G..   [198]
CRNP9           .................C................................   [199]
```

```
CRNP10      .................C.....C..........................  [199]
CRNP12      ............T....C.........T........AC.........G..  [198]
AKZN6       .................C................................  [199]
AKZN7       .................C................................  [199]
AKZN8       .................C................................  [199]
AKZN10      .................C....................G...........  [199]
GONCOMP01   .................C....................G...........  [199]
AKZN13      .................C.........T......................  [199]
AKZN14      .................C................................  [199]
AKZN16      .................C................................  [199]
NGNY14      .................C..........................C.....  [199]
NGNY15      .................C..........................C.....  [199]
GGNP951     .................C................................  [199]
GGNP954     .................C....................G...........  [199]
GGNP955     .................C................................  [199]
GGNP21      .................C................................  [199]
GGNP23      .................C................................  [199]
GGNP27      .................C....................G...........  [199]
BYM1        .................C................................  [199]
BYM3        .................C.........T........AC............  [198]
BYM4        .................C................................  [199]
BYM5        .................C................................  [199]
MTCM2       .................C................................  [199]
MTCM4       .................C................................  [199]
MTCM7       .................C................................  [199]
MTCM8       .................C..................A.............  [199]
MTCM11      .................C................................  [199]
MTCM12      .................C................................  [199]
MTCM15      .................C................................  [199]
MTCM16      .................C................................  [199]
MTCM21      .................C................................  [199]
Mosse1      ..........T......C................................  [199]
Mosse2      ..........T......C................................  [199]
Mosse5      .................C...............A................  [199]
Mosse8      .................C................................  [199]
Mosse12     .................C................................  [199]
Mosse13     .................C................................  [199]
Mosse14     .................C................................  [199]
Mosse15     .................C................................  [199]
Mosse21     ..........T......C................................  [199]
Man1        ............T....C.........T........AC........TG..  [198]
Man3        ............T....C.........T........AC........TG..  [198]
Man6        .................C................................  [199]
Man7        .................C................................  [199]
Man8        ............T....C.........T........AC........TG..  [198]
Man10       .................C................................  [199]
Man11       ..........T......C................................  [199]
Man12       .................C................................  [199]
Man15       .................C................................  [199]
Man16       .................C................................  [199]
DEFR1       ............T....C.........T........AC.......G.G..  [198]
DEFR2       ............T....C.........T........AC.......G.G..  [198]
DEFR3       ............T....C.........T........AC.........G..  [198]
DEFR6       .................C......T...........AC........TG..  [199]
```

```
DEFR7         ............T....C...T.....T........AC......G..G..   [198]
DEFR8         ............T....C.........T........AC.........G..   [197]
DEFR9         ............T....C.........T........AC.......G.G..   [198]
Campo1        ............T....C.........T........AC.........G..   [198]
Campo2        .................C.....G...T........AC.........G..   [198]
Campo6        .................C.....G...T........AC.........G..   [198]
GONCOMP02     ............T....C.........T........AC.........G..   [198]
Campo9        ..........T.T....C.........T........AC............   [198]
Campo13       ............T....C.........T........ACG...........   [198]
Campo14       ............T....C.........T........AC.........G..   [198]
Campo15       ............T....C.........T........ACG...........   [198]
DBR2          .................C...................C........TG..   [199]
DBR3          .................C...................C........TG..   [199]
DBR5          .................C..................AC........TG..   [199]
GONCOMP03     .................C..................AC........TG..   [199]
DBR7          .................C................................   [198]
DBR10         ..........T.T...GC.........T........AC.........G..   [198]
DBR12         ............T....C...T.....T........AC......G..G..   [198]
DBR13         .................C..................AC........TG..   [199]
ISEFR1        ...............T.C........T.........A.............   [199]
ISEFR3        ...............G.C........T.........A.............   [199]
Gimble        .................C.........T........AC............   [198]
MORCOMP01     .................C.........T........AC............   [198]
MORCOMP02     .................C.........T........AC............   [198]
MORCOMP03     .................C.........T........AC............   [198]
WL_3          .................C.........T........AC............   [198]
MORCOMP04     .................C.........T........AC............   [198]
George        .................C.........T........AC............   [198]
Pts_43        ..........T......C.........T........AC............   [198]
MORCOMP05     ..........T.T....C.........T........AC.........G..   [198]
Gemini        ............T....C.........T........AC.........G..   [198]
Ivindo        ............T....C.........T........AC......G..G..   [198]
Makokou       ............T....C.........T........AC......G..G..   [198]
Mopia         ............T....C.........T........AC.........G..   [198]
Mouilla       ............T....C.........T........AC.........G..   [198]
Ponia         ..........T.T....C.........T........AC.........G..   [198]
Ptt_36        .................C.G................AC........TG..   [200]
Ptt_37        .................C.....G...T........AC........TG..   [198]
Berthe        .................C.........T.........C........TG..   [199]
Milla         .........?.?.....C...?????..........AC........TG..   [200]
C1            ............T....C.........T........AC.........G..   [199]
Ptv_4         .................C........T.T....?..A.............   [198]
Ptv_5         .................C........T.......................   [199]
Ptv_15        .................C........T.................G.....   [199]
C2            .................C........T.T....A..A.............   [199]
C3            .................C.......T........................   [198]
GAGCOMP01     .................C........T.T....A..A.............   [198]
GAGCOMP02     .................C........T.......................   [199]
Ulsse8430     .................C........T.......................   [199]
RICCInc58     .................C...............A................   [198]
GAGCOMP03     .................C.......T........................   [199]
13_p84339     .................C........T.......................   [199]
12_84303      .................?...............A................   [198]
GAGCOMP04     .................C........T.........A.............   [199]
```

```
GAGCOMP05       .................C........T.T....A..A.............   [198]
Ali             .................C........T.......................   [199]
GAGCOMP6        ....T............C........T.................G.....   [199]
Darwin          .................C........T.........A.............   [199]
Dilly           ...........???...C........T.T....A..A.............   [198]
Hera            .................C.......T.............?..........   [199]
Loukoum         ............?....C........T.T....A..A.............   [198]
Ondine          .................C...............A................   [198]
Perla           .................C........T.T....A..A.............   [198]
Rousseau        .................C........T.........A.............   [199]
Salome          .................C................................   [199]
Venus           .................C.......T........................   [199]
Xeres           ............??...C........T.........C.?...........   [199]
B3              .................C........T.T....A..A.............   [198]
F18             .................C........T.T....A..A.............   [198]
K22             .................C?.......T.T....A..A.............   [198]
M19             .................C........T.T....A..A.............   [198]
M18             .................C........T.T....A..A.............   [198]
Solo10          .................C........T.T....A..A.............   [199]
Solo13          .................C........T.......................   [199]
Wongo1          .?..?.........?..?.?...?..T...?...................   [199]
Wongo7          .................C........T.......................   [199]
Wongo5          .................C...............A.........?......   [198]
Pts001.         .................C.........T........AC............   [198]
Pts002          .................C.........T........AC............   [198]
Pts003          .................C.........T........AC............   [198]
Pts005.         .................C.........T........AC.......?....   [198]
Pts006.         .................C.........T........AC............   [198]
Pts007.         .................C.........T........AC............   [198]
Pts008.         ..........T......C.........T........AC............   [198]
Pts009.         .................C.........T........AC............   [198]
Pts010.         .................C.........T........AC............   [198]
Pts011.         .................C.........T........AC............   [198]
Pts012.         ..........T.T....C.........T........AC............   [198]
Pts013.         ..........T......C.........T........AC............   [198]
Pts014.         ..........T......C.........T........AC............   [198]
Pts015.         .................C.........T........AC............   [198]
Pts016.         .................C.........T........AC............   [198]
Pts017.         .................C.........T........AC............   [198]
Pts018.         .................C.........T........AC............   [198]
Pts019.         .................C........TT........AC............   [198]
Pts020.         .................C........TT........AC............   [198]
Pts021.         .................C........TT........AC............   [198]
Pts022.         .................C.........T........AC............   [198]
Pts023.         .................C.........T........AC............   [198]
Pts024.         .................C.........T........AC............   [198]
Pts025.         .................C.........T........AC............   [198]
Pts026.         .................C.........T........AC............   [198]
Pts027.         .................C.........T........AC............   [198]
Pts028.         .................C.........T........AC............   [198]
Pts029.         .................C.........T........AC............   [198]
Pts030.         .................C.........T........AC............   [198]
Pts031.         ......?.........?C.........T........AC............   [198]
Pts032.         .................C.........T........AC............   [198]
```

```
Pts033.     .................C.........T..G.....AC............   [198]
Pts034.     .................C.........T........AC............   [198]
Pts035.     ..........T......C.........T........AC............   [198]
Pts036.     .................C.........T........AC............   [198]
Pts037.     .................C.........T........AC............   [198]
Pts038.     .................C.........T........AC............   [198]
Pts039.     .................C.........T........AC............   [198]
Pts040.     .................C.........T........AC............   [198]
Pts041.     ......?..........C.........T........AC............   [198]
Pts043.     .................C.........T........AC............   [198]
Pts044.     .................C.........T........AC............   [198]
Pts045.     .................C.....G...T........AC............   [198]
Pts046.     .................C.........T........AC............   [198]
Pts047.     ..........T......C.........T........AC............   [198]
Pts048.     ..........T.T....C.........T........AC............   [198]
Pts049.     .................C.........T........AC............   [198]
Pts050.     .................C.........T........AC............   [198]
Pts051.     .................C.........T........AC............   [198]
Pts052.     .................C.........T........AC............   [198]
Pts053.     .................C.........T........AC............   [198]
Pts054.     .................C.........T........AC............   [198]
Pts055.     .................C.........T........AC............   [198]
Pts056.     .................C.........T........AC............   [198]
Pts057.     .................C.........T........AC............   [198]
Pts058.     .................C.........T........AC............   [198]
Pts059.     .................C.........T........AC............   [198]
Pts060.     .................C.........T........AC............   [198]
Pts061.     .................C.........T........AC............   [198]
Pts062.     .................C.........T........AC............   [198]
Pts063.     .................C.........T........AC............   [198]
Pts065.     ..........T.T....C.........T........AC............   [198]
Pts066.     .................C.........T........AC............   [198]
Pts067.     .................C...T.....T........AC............   [198]
Pts068.     .................C.........T........AC............   [198]
Pts069.     .................C.........T........AC............   [198]
Pts070.     .................C.........T........AC............   [198]
Pts071.     .................C.........T........AC............   [198]
Pts072.     ..........T......C.........T........AC............   [198]
Pts073.     .................C.........T........AC............   [198]
Pts074.     ..........T......C.........T........AC............   [198]
Pts075.     ..........T......C.........T........AC............   [198]
Pts076.     .................C.........T........AC............   [198]
Pts077.     .................C.........T........AC............   [198]
Pts078.     .................C.........T........AC............   [198]
Pts079.     ..........T......C.........T........AC............   [198]
Pts080.     ..........T......C.........T........AC............   [198]
Pts081.     ..........T......C..................AC............   [198]
Pts082.     ..........T......C.........T........AC............   [198]
Pts083.     ..........T......C..................AC............   [198]
Pts084.     ..........T......C..................AC............   [198]
Pts085.     .................C...T.....T........AC............   [198]
Pts086.     .................C.........T........AC............   [198]
GOLCOMP01   .................C.........T.......TAC............   [198]
Pts089.     .................C.........T.......TAC............   [198]
```

```
Pts090.       .................C.........T.......TAC............   [198]
Pts091.       .................C.........T........AC............   [198]
Pts092.       .................C.........T........AC............   [198]
Pts093.       .................C.........T........AC............   [198]
Pts094.       .................C.........T........AC............   [198]
Pts095.       .................C.........T........AC............   [198]
Pts096.       .................C.........T........AC............   [198]
Pts097.       .................C.........T........AC............   [198]
Pts098.       .................C.........T........AC............   [198]
Pts099.       .................C.........T........AC............   [198]
Pts100.       .................C...T.....T........AC............   [198]
Pts101.       .................C.........T........AC............   [198]
Pts102.       .................C...T.....T........AC............   [198]
Pts103.       .................C.........T........AC............   [198]
Pts104.       .................C.........T........AC............   [198]
Pts105.       .................C.........T........AC............   [198]
Pts106.       .................C.........T........AC............   [198]
Pts107.       .................C.........T........AC............   [198]
Pts108.       .................C.........T........AC............   [198]
Pts109.       .................C.........T........AC............   [198]
Pts110.       .................C.........T........AC............   [198]
Pts111.       .................C.........T........AC............   [198]
Pts112.       .................C...T.....T........AC............   [198]
Pts113.       .................C.........T........AC............   [198]
Pts114.       .................C.........T........AC............   [198]
Pts115.       .................C.........T........AC............   [198]
Pts116.       .................C.........T........AC............   [198]
Pts117.       .................C.........T........ACG...........   [198]
Pts118.       .................C.........T........AC............   [198]
Pts119.       .................C.........T........AC............   [198]
Pts120.       .................C.........T........AC............   [198]
Pts121.       .................C.....?...T........AC.......?....   [198]
Pts122.       .................C.........T........AC............   [198]
Pts123.       .................C.........T........AC............   [198]

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

Human         ACTCCAAAGCCACCCCTCACCCACTAGGATATCAACAAACCTACCCGCCC   [249]
H45           ACCCCAAAGCCACCCCTCACCCACTAGAATATCAACAAACCTACCCATCC   [249]
Ppaniscus     CCCCCAAAGACACTCCCCCCCCACCCCGATACCAACAAACCTGACAGTCC   [248]
P_1Kp         CCCCCAAAGACACTCCCCCCCCACCCCGATACCAACAAACCTGACAGTCC   [248]

CRNP2         ACTCCAACGACATTCCTACCCCACCCCGATACCAACAAACCTATTCCTCC   [249]
CRNP3         ............C..............A..............G.C.T...   [249]
CRNP6         .......A....C....C.........................CA.TC..   [248]
CRNP9         ..................................................   [249]
CRNP10        ............C..............A..............G.C.T...   [249]
CRNP12        .......A....C....CT........................CA.TC..   [248]
AKZN6         ..................................................   [249]
AKZN7         ............................................C.....   [249]
AKZN8         ............................................C.....   [249]
AKZN10        ............CC..............................C.....   [249]
GONCOMP01     ............CC..............................C.....   [249]
```

```
AKZN13       ........................T.........................   [249]
AKZN14       ..................T.....................T.........   [249]
AKZN16       ............................................C.....   [249]
NGNY14       .................C................................   [249]
NGNY15       .................C................................   [249]
GGNP951      ..................T.....................T.........   [249]
GGNP954      ............?C..............................C.....   [249]
GGNP955      ..................T.....................T.........   [249]
GGNP21       ............................................C.....   [249]
GGNP23       ............CC...C..........................C...T.   [249]
GGNP27       ............CC..............................C.....   [249]
BYM1         ...........................................C......   [249]
BYM3         .......A.........C.............T.....G......CTTC..   [248]
BYM4         ............C.....................................   [249]
BYM5         ..................................................   [249]
MTCM2        ............CC..............................C.....   [249]
MTCM4        .......T.G..CC...C..........................C...T.   [249]
MTCM7        .......T.G..CC...C.......?..................C.T...   [249]
MTCM8        ..................................................   [249]
MTCM11       .......T.G..CC...C..........................C...T.   [249]
MTCM12       ...........................................C......   [249]
MTCM15       ............CC..............................C.....   [249]
MTCM16       ..................................................   [249]
MTCM21       .......T....CC...C..........................C...T.   [249]
Mosse1       ........................T............G..T.........   [249]
Mosse2       ........................T............G..T.........   [249]
Mosse5       ..................T...............................   [249]
Mosse8       ............................................C.....   [249]
Mosse12      .......T.G..CC...C..........................C...T.   [249]
Mosse13      ..................................................   [249]
Mosse14      .......T.G..CC...C..........................C...T.   [249]
Mosse15      ...........................................C......   [249]
Mosse21      ........................T............G..T.........   [249]
Man1         .......A....C....C.........................CA.TC..   [248]
Man3         .......A....C....C.........................CA.TC..   [248]
Man6         .......T....CC...C..........................C...T.   [249]
Man7         ...........................................C......   [249]
Man8         .......A....C....C.........................CA.TC..   [248]
Man10        ..................................................   [249]
Man11        ..................................................   [249]
Man12        ...........................................C......   [249]
Man15        ..................................................   [249]
Man16        ..................................................   [249]
DEFR1        .......A....C....C.........................CA.TC..   [248]
DEFR2        .......A....C....C.........................CA.TC..   [248]
DEFR3        .......A....C....CT........................CA.TC..   [248]
DEFR6        .......A....CC...C.............................CTT   [249]
DEFR7        .......A....CC...C.........................CA.TC..   [248]
DEFR8        .......A....C....C.........................CA.TC..   [247]
DEFR9        .......A....C....C.........................CA.TC..   [248]
Campo1       .......A....C....CT........................CA.TC..   [248]
Campo2       .......A....C....C...................G....GCA..C..   [248]
Campo6       .......A....C....C...................G....GCA..C..   [248]
GONCOMP02    .......A....C....CT........................CA.TC..   [248]
```

```
Campo9       .......A..T.C....C..........................A..CT.   [248]
Campo13      .......A....C....C.........................CA.TC..   [248]
Campo14      .......A....C....CT........................CA.TC..   [248]
Campo15      .......A....C....C.........................CA.TC..   [248]
DBR2         .......A....CC...CT.........................A...T.   [249]
DBR3         .......A....CC...CT.........................A...T.   [249]
DBR5         .......A....CC...CT.....T...................A...TT   [249]
GONCOMP03    .......A....CC...CT.........................A...TT   [249]
DBR7         ............CC..............................C.....   [248]
DBR10        .......A....C....C........T................CA.TC..   [248]
DBR12        .......A....C....C.........................CA.TC..   [248]
DBR13        .......A....CC...C.............................CT.   [249]
ISEFR1       .................C.........A................C.....   [249]
ISEFR3       .................C.........A................C.....   [249]
Gimble       .......A.........C...................G.....CA.TC..   [248]
MORCOMP01    .......A....C....C..........................A..CT.   [248]
MORCOMP02    .......A.........C......T..A...T............C.TCT.   [248]
MORCOMP03    .......A....C....C........T.................A.TCT.   [248]
WL_3         .......A.........C.........AG..T............C.TC..   [248]
MORCOMP04    .......A.........C.............T............C.TCT.   [248]
George       .......A.........C...................G......A.TCT.   [248]
Pts_43       .......A....C....C..........T.............G.A.TC..   [248]
MORCOMP05    .......A.T.......C.T......T.................A.TC..   [248]
Gemini       .......A....C....C.........................CA.TC..   [248]
Ivindo       .......A....C....C.........................CA.TC..   [248]
Makokou      .......A....C....C.........................CA.TC..   [248]
Mopia        .......A....C....C.........................CA.TC..   [248]
Mouilla      .......A....C....C.........................CA.TC..   [248]
Ponia        .......A.G.......C......................T..CA.TC..   [248]
Ptt_36       .......A....CC...C..........................A..C..   [250]
Ptt_37       .......A....C....C....T..............G....GCA..C..   [248]
Berthe       .......A....CCT..CT.........................A...TT   [249]
Milia        .......A....C....??.........................A..C..   [250]
C1           .......A....C....C.........................CA.TC..   [249]
Ptv_4        ............C.T..C.........................CC..C..   [248]
Ptv_5        ............CC...C.............T.....G.....CC..C.-   [248]
Ptv_15       ............CC...C.............T.....G.....CC.....   [249]
C2           ............C.T..C.........................CC..C..   [249]
C3           .T..........CC...C...................G......CT.C..   [248]
GAGCOMP01    ............C.T..C.........................CC..C..   [248]
GAGCOMP02    ............CC...C.............T.....G.....CC..C..   [249]
Ulsse8430    ............CC...C.............T.....G.....CC..C..   [249]
RICCInc58    ............C.T..C........TA...............CC.TC..   [248]
GAGCOMP03    .T..........CC...C...................G.....CCT.C..   [249]
13_p84339    ............CC...C.............T...........CC..C..   [249]
12_84303     ............C.T..C........?A...............CC.TC..   [248]
GAGCOMP04    ............CC...C.............T.....G.....CC..C.-   [248]
GAGCOMP05    ............CCT..C.........................CC..C.-   [247]
Ali          ........?...CC...C.............T.....G.....CC..C.-   [248]
GAGCOMP6     ............CC...C.............T.....G.....CC..C.-   [248]
Darwin       ............CC...C.............T.....G.....CC..C.-   [248]
Dilly        ............C....C.........................CC..C..   [248]
Hera         .T..........CC...C...................G......CT.C..   [249]
Loukoum      ............CCT..C.........................CC..C.-   [247]
```

```
Ondine      ............C.T..C........TA...............CC.TC..   [248]
Perla       ............C.T..C.........................CC..C..   [248]
Rousseau    ............CC...C.........A...T.....G.....CC..C.-   [248]
Salome      .T..........CC...C...................G......C..C..   [249]
Venus       .T..........CC...C...................G......C..C.-   [248]
Xeres       ............CC...C.............T.....G.....CC..C..   [249]
B3          ............CCT..C.........A...............CC..C..   [248]
F18         ............C....C.........................CC..C..   [248]
K22         ............CCT..C.........................CC..C..   [248]
M19         ............C.T..C.........................CC..C..   [248]
M18         ............CCT..C.........................CC..C.-   [247]
Solo10      ....?.......CCT..C.........................CC..C.-   [248]
Solo13      ............CC...C........T....T.....G.??..CC..C.-   [248]
Wongo1      .......?....CC...C.............T.....G.....CC..C..   [249]
Wongo7      ............CCT..C.............T.....G.....CC..C..   [249]
Wongo5      ............C.T..C........TA.....?.........CC.TC..   [248]
Pts001.     .......A.........C..........................A.TC..   [248]
Pts002      .......A.........C..........................A.TC..   [248]
Pts003      .......A.........C..........................A.TC..   [248]
Pts005.     .......A....C....C..........................A.TCT.   [248]
Pts006.     .......A....C....C..........................A.TCT.   [248]
Pts007.     .......A....C....C.T........................A.TCT.   [248]
Pts008.     .......A....C....C..........................A.TCT.   [248]
Pts009.     .......A....C....C...................G......A.TCT.   [248]
Pts010.     .......A....C....C..........................A.TCT.   [248]
Pts011.     .......A.........C........T..........G......A.TCT.   [248]
Pts012.     .......A.........C........T.................A.T...   [248]
Pts013.     .......A.........C........T.................A.T.T.   [248]
Pts014.     .......A.........C........T.................A.T.T.   [248]
Pts015.     .......A.........C...................G......A..CT.   [248]
Pts016.     .......A.........C..........................A..CT.   [248]
Pts017.     .......A....C....C..........................A..C..   [248]
Pts018.     .......A....C....C..........................A..CT.   [248]
Pts019.     .......A.........C........T.................A.TCT.   [248]
Pts020.     .......A.........C........T..........G......A.TCT.   [248]
Pts021.     .......A.........C........T.................A.TCT.   [248]
Pts022.     .......A....C....C........T.................A.TCT.   [248]
Pts023.     .......A....C....C........T.................A.TCT.   [248]
Pts024.     .......A....C....C........T.................A.TCT.   [248]
Pts025.     .......A....C....C........T.................A.TCT.   [248]
Pts026.     .......A....C....C........T.................A.TCT.   [248]
Pts027.     .......A....C....C........T.................A.TCT.   [248]
Pts028.     .......A.........C.........AG..T.....G......C.TC..   [248]
Pts029.     .......A.........C.........AG..T............C.TC..   [248]
Pts030.     .......A.........C.........AG..T............C.TC..   [248]
Pts031.     .......?.........C.........AG..T......?.....C.TC..   [248]
Pts032.     .......A.........C......T..A...T............C.TCT.   [248]
Pts033.     .......A.........C.........A...T............C.TCT.   [248]
Pts034.     .......A.........C.........A...T............C.TCT.   [248]
Pts035.     .......A.........C.............T............C.TCT.   [248]
Pts036.     .T.....A.........C.............T............C.TCT.   [248]
Pts037.     .T.....A.........C.............T............C.TCT.   [248]
Pts038.     .T.....A.........C.............T............C.TCT.   [248]
Pts039.     .T.....A.........C.............T............C.TCT.   [248]
```

```
Pts040.     .......A.........C.......T...........G......A.TCT.   [248]
Pts041.     .......A.........C.............T............C.TCT.   [248]
Pts043.     .......A.........C.............T............C.TCT.   [248]
Pts044.     .......A....C....C.............T............C.TCT.   [248]
Pts045.     .......A.........C..........................A.TCT.   [248]
Pts046.     .......A....C....C.............T............A.TCT.   [248]
Pts047.     .......A....C....C.............T............A.TCT.   [248]
Pts048.     .......A.........C...................G......A.TCT.   [248]
Pts049.     .......A.........C...................G......A.TCT.   [248]
Pts050.     .......A.........C...................G......A.TC..   [248]
Pts051.     .......A.........C...................G......A.TC..   [248]
Pts052.     ......GA.........C...................G....G.A.TC..   [248]
Pts053.     ......GA.........C...................G....G.A.TC..   [248]
Pts054.     .......A.........C...................G....G.A.TC..   [248]
Pts055.     .......A.........C...................G....G.A.TCT.   [248]
Pts056.     .......A.........C...................G....G.A.TCT.   [248]
Pts057.     .......A.....C...C...................G....G.A.TC..   [248]
Pts058.     .......A.....C...C........................G.A.TC..   [248]
Pts059.     .......A.........C...................G....G.A.TC..   [248]
Pts060.     .......A.........C...................G....G.A.TCT.   [248]
Pts061.     .......A.........C...................G....G.A.TC..   [248]
Pts062.     .......A.........C...................G....G.A.TC..   [248]
Pts063.     .......A.........C...................G....G.A.TC..   [248]
Pts065.     .......A.........C...................G....G.A.TC..   [248]
Pts066.     .......A.........C...................G......A.TC..   [248]
Pts067.     .......A.........C.T.................G......A.TCT.   [248]
Pts068.     .......A....C....C..........................A.TC..   [248]
Pts069.     .......A........CC..........................A.TC..   [248]
Pts070.     .......A.........C..........................A.T...   [248]
Pts071.     .......A.........T...................G......A..CT.   [248]
Pts072.     .......A.........T...................G......A..CT.   [248]
Pts073.     .......A.........C...................G......A.TCT.   [248]
Pts074.     .......A.........C.............T.....G......A.TCT.   [248]
Pts075.     .......A.........C...................G......A.TCT.   [248]
Pts076.     .......A.........C...................G......A.TCT.   [248]
Pts077.     .......A.........C...................G......A.TCT.   [248]
Pts078.     .......A.........C...................G......A.TCT.   [248]
Pts079.     .......A.........C..........................A.TC..   [248]
Pts080.     .......A.........C..........................A.TCT.   [248]
Pts081.     .......A.........C..........................A.TCT.   [248]
Pts082.     .......A.........C..........................A.TCT.   [248]
Pts083.     .......A.........C..........................A.TCT.   [248]
Pts084.     .......A.........C..........................A.TCT.   [248]
Pts085.     .......A.........C...................G......A.TCT.   [248]
Pts086.     .......A.........C.........................CA.TCT.   [248]
GOLCOMP01   .......A.........C.T........................A.TCT.   [248]
Pts089.     .......A.........C..........................A.TCT.   [248]
Pts090.     .......A.........C..........................A.TCT.   [248]
Pts091.     .......A......T..C...................G......A.TCT.   [248]
Pts092.     .......A......T..C...................G......A.TCT.   [248]
Pts093.     .......A......T..C...................G......A.TCT.   [248]
Pts094.     .......A......T..C...................G......A.TCT.   [248]
Pts095.     .......A......T..C...................G......A.TCT.   [248]
Pts096.     .......A......T..C...................G......A.TCT.   [248]
```

```
Pts097.         .......A....A....C...................G......A.TCT.    [248]
Pts098.         .......A....A....C...................G......A.TCT.    [248]
Pts099.         .......A....A....C...................G......A.TCT.    [248]
Pts100.         .......A.........C...................G......A.TCT.    [248]
Pts101.         .......A.........C...................G......A.TCT.    [248]
Pts102.         .......A.........C...................G......A.TCT.    [248]
Pts103.         .......A.........C...................G......A.TCT.    [248]
Pts104.         .......A.........C.......T...........G......A.TCT.    [248]
Pts105.         .......A.........C...................G......A.TCT.    [248]
Pts106.         .......A.........C...................G......A.TCT.    [248]
Pts107.         .......A.........C...................G......A.TCT.    [248]
Pts108.         .......A.........C...................G......A.TCT.    [248]
Pts109.         .......A.........C..........................A.TCT.    [248]
Pts110.         .......A.........C..........................A.TCT.    [248]
Pts111.         .......A.........C...................G......A.TCT.    [248]
Pts112.         .......A.........C...................G......A.TCT.    [248]
Pts113.         .......A.........C...................G......A.TCT.    [248]
Pts114.         .T.....A.........C.............T............C.TCT.    [248]
Pts115.         .......A.........C...................G......A.TCT.    [248]
Pts116.         .......A.........C...................G......A.TCT.    [248]
Pts117.         .......A.........C...................G......A.TC..    [248]
Pts118.         .......A.........C...................G......A.TC..    [248]
Pts119.         .......A.........C...................G......A.TC..    [248]
Pts120.         .......A.........C...................G......A.TC..    [248]
Pts121.         .......A.........C....?..............G.....CA.TC..    [248]
Pts122.         .T.....A.........C...................G....G.A.TC..    [248]
Pts123.         .......A.........C........T..........G......A.TC..    [248]

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

Human           TTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAG    [299]
H45             TTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAG    [299]
Ppaniscus       TTAACAGTACATAGCACATACAATTATATACCGTACATAGCACATTACAG    [298]
P_1Kp           TTAACAGTACATAGCACATACAATTATATACCGTACATAGCACATTACAG    [298]

CRNP2           TTAACAGGACATAGCACATACAATCATACACCGTACATAGCACATTACAG    [299]
CRNP3           .......A..............G...C.......................    [299]
CRNP6           .......T...............C.G......AG....G...........    [298]
CRNP9           ...G......................C.......................    [299]
CRNP10          .......A..............G...C......................A    [299]
CRNP12          .......T...............C.G...G..AG....G...........    [298]
AKZN6           ..........................C.......................    [299]
AKZN7           ..................................................    [299]
AKZN8           .........................G........................    [299]
AKZN10          ......................G...........................    [299]
GONCOMP01       ......................G...........................    [299]
AKZN13          ..........................C.......................    [299]
AKZN14          ..........................C.......................    [299]
AKZN16          .........................G........................    [299]
NGNY14          ..........................C.......................    [299]
NGNY15          ..........................C.......................    [299]
GGNP951         ..........................C.......................    [299]
GGNP954         ......................G...........................    [299]
```

```
GGNP955     ..........................C.......................   [299]
GGNP21      .........................G........................   [299]
GGNP23      ......A........G......G...C.T.....................   [299]
GGNP27      ......................G...........................   [299]
BYM1        ..........................C.......................   [299]
BYM3        .......T......T.......TC.G....T.AC................   [298]
BYM4        ..........................C.......................   [299]
BYM5        ..........................C.......................   [299]
MTCM2       ......................G...........................   [299]
MTCM4       ......A...............G...C.T..................T..   [299]
MTCM7       ......................G...C.T.....................   [299]
MTCM8       ..................................................   [299]
MTCM11      ......A...............G...C.T..................T..   [299]
MTCM12      ..........................C.......................   [299]
MTCM15      ......................G...........................   [299]
MTCM16      ..........................C.......................   [299]
MTCM21      ......A...............G...C.T.....................   [299]
Mosse1      ....T.....................C.......................   [299]
Mosse2      ....T.....................C.......................   [299]
Mosse5      ..........................C.......................   [299]
Mosse8      .........................G........................   [299]
Mosse12     ......A...............G...C.T..................T..   [299]
Mosse13     ..................................................   [299]
Mosse14     ......A...............G...C.T..................T..   [299]
Mosse15     ..........................C.......................   [299]
Mosse21     ....T.....................C.......................   [299]
Man1        .......T...............C.GC.....AG................   [298]
Man3        .......T...............C.GC.....AG................   [298]
Man6        ...............G......G...C.T.....................   [299]
Man7        ......................G...C.......................   [299]
Man8        .......T...............C.GC.....AG................   [298]
Man10       .......A..................C.......................   [299]
Man11       ..........................C.......................   [299]
Man12       ......................G...C.......................   [299]
Man15       ..................................................   [299]
Man16       ..........................C.......................   [299]
DEFR1       .......T...............C.G......AG....G...........   [298]
DEFR2       .......TC..............C.G......AG....G...........   [298]
DEFR3       .......T..............GC.G...?..AG....G...........   [298]
DEFR6       .......T..............GC.G....T..C................   [299]
DEFR7       .......T...............C.GC.....AG.............T..   [298]
DEFR8       .......T...............C.GC.....AG.............T..   [297]
DEFR9       .......T...............C.G......AG....G...........   [298]
Campo1      .......T...............C.G...G..AG....G...........   [298]
Campo2      ....T..T...............C.GC.....AG....G...........   [298]
Campo6      ....T..T...............C.GC.....AG....G...........   [298]
GONCOMP02   .......T...............C.G...G..AG....G...........   [298]
Campo9      .......T......T........C.G......A.................   [298]
Campo13     .......T...............C.G......AG................   [298]
Campo14     .......T...............C.G...G..AG....G...........   [298]
Campo15     .......T...............C.G......AG................   [298]
DBR2        .......T......T........C.G....T..C................   [299]
DBR3        ..............T........C.G....T..C................   [299]
DBR5        .......T......T.....T..C.G....T..C................   [299]
```

```
GONCOMP03     .......T......T.....T..C.G....T..C................    [299]
DBR7          ......................G...........................    [298]
DBR10         .......T...............C.GC.....AG................    [298]
DBR12         .......T...............C.GC.....AG.............T..    [298]
DBR13         .......T..............GC.G....T..C................    [299]
ISEFR1        .............AT.......G.T......................G...    [299]
ISEFR3        .............AT.......G.T......................G...    [299]
Gimble        .......T......T........C.G......A.................    [298]
MORCOMP01     .......T...............C.G......A.................    [298]
MORCOMP02     .......T......T........C.G......A..............T..    [298]
MORCOMP03     .......T...............C.G......A.................    [298]
WL_3          .......T......T........C.G......A..............T..    [298]
MORCOMP04     .......T......T........C.G......A.................    [298]
George        .......T......T........C.G......A.................    [298]
Pts_43        ..TT.T.T......T........?.G.....?A.................    [298]
MORCOMP05     .......T......T........C.GC...T.AC................    [298]
Gemini        .......T...............C.GC.....AG.............T..    [298]
Ivindo        .......T...............C.GC.....AG................    [298]
Makokou       .......T...............C.GC.....AG................    [298]
Mopia         .......T...............C.GC.....AG....G...........    [298]
Mouilla       .......T......T........C.GC.....AG.............T..    [298]
Ponia         .......C...............C.GC.....AG....G...........    [298]
Ptt_36        .......T......T........C.G....T..C................    [300]
Ptt_37        ....T..T...............C.GC.....AG....G...........    [298]
Berthe        .......T......T........C.G....T..C................    [299]
Milla         .......CG.....T........C.G....T..............CG...    [300]
C1            .......T...............C.GC.....AG................    [299]
Ptv_4         ..G...AA......T...........C.......................    [298]
Ptv_5         ..G....A......T...................................    [298]
Ptv_15        C.G....A......T...................................    [299]
C2            ..G...AA......T...........C.......................    [299]
C3            ..G....A......T........C..........................    [298]
GAGCOMP01     ..G...AA......T...........C.......................    [298]
GAGCOMP02     ..G....A......T...................................    [299]
Ulsse8430     ..G....A......T...................................    [299]
RICCInc58     ..G...AA.....AT.........T.C.......................    [298]
GAGCOMP03     ..G....A......T........C..........................    [299]
13_p84339     ..G....?......T...................................    [299]
12_84303      ..G...AA.....AT...........C.......................    [298]
GAGCOMP04     ..G....A......T...................................    [298]
GAGCOMP05     ..G...AA......T...........C.......................    [297]
Ali           ..G....A......T...................................    [298]
GAGCOMP6      ..G....A......T...................................    [298]
Darwin        ..G....A......T.................................G.    [298]
Dilly         ..G...A.......T...........C.........G.............    [298]
Hera          ..G....A......T........C..........................    [299]
Loukoum       ..G...AA......T...........C.......................    [297]
Ondine        ..G...AA.....AT.........T.C.......................    [298]
Perla         ..G...AA......T...........C.......................    [298]
Rousseau      ..G...........T...................................    [298]
Salome        ..G....A......T........C....T.....................    [299]
Venus         ..G...........T........C..........................    [298]
Xeres         ..G....A......T...................................    [299]
B3            ..G...AA......T...........C.......................    [298]
```

```
F18            ..G...A.......T...........C.......................   [298]
K22            ..G...AA......T...........C.......................   [298]
M19            ..G...AA......T...........C.......................   [298]
M18            ..G...AA......T...................................   [297]
Solo10         ..G...AA......T...........C.......................   [298]
Solo13         ..G.T..A.....AT..................................C   [298]
Wongo1         ..G.T..A.....AT..............?...................C   [299]
Wongo7         ..G.T..A.....AT..................................C   [299]
Wongo5         ..G...AA..?..AT...........C.......?...............   [298]
Pts001.        .......T......T..........G.....TA................A   [298]
Pts002         .......T......T..........G.....TA.................   [298]
Pts003         .......T......T..........G.....TA................A   [298]
Pts005.        .......T...............C.G....T.A.................   [298]
Pts006.        .......T..............GC.G....T.A.................   [298]
Pts007.        .......T...............C.G....T.A.................   [298]
Pts008.        .......T...............C.G....T.A.................   [298]
Pts009.        .......T...............C.G....T.A.................   [298]
Pts010.        .......T...............C.G....T.A.................   [298]
Pts011.        .......T......T....T...C.G......A..............G..   [298]
Pts012.        ...G...T......T........C.G......A.................   [298]
Pts013.        .......T......T........C.G......A.................   [298]
Pts014.        .......T......T........C.G......A.................   [298]
Pts015.        .......T...............C.G......A.................   [298]
Pts016.        .......T...............C.G......A.................   [298]
Pts017.        .......T...............C.G......A.................   [298]
Pts018.        .......T...............C.G......A.................   [298]
Pts019.        .......T...............C.G......A.................   [298]
Pts020.        .......T...............C.G....T.A.................   [298]
Pts021.        .......T...............C.G....T.A.................   [298]
Pts022.        .......T...............C.G.....TA.................   [298]
Pts023.        .......T...............C.G......A.................   [298]
Pts024.        .......T...............C.G......A.................   [298]
Pts025.        .......T...............C.G......A.................   [298]
Pts026.        .......T...............C.G......A.............G...   [298]
Pts027.        .......T...............C.G......A.............G...   [298]
Pts028.        .......T......T........C.G......A..............T..   [298]
Pts029.        .......T......T........C.G......A..............T..   [298]
Pts030.        .......T......T........C.G......A..............T..   [298]
Pts031.        .......T......T........C.G......A...?.........?T..   [298]
Pts032.        .......T......T........C.G......A..............T..   [298]
Pts033.        .......T......T........C.G......A..............T..   [298]
Pts034.        .......T......T........C.G......A..............T..   [298]
Pts035.        .......T......T........C.G......A.................   [298]
Pts036.        .......T......T........C.G......A.................   [298]
Pts037.        .......T......T........C.G......A.................   [296]
Pts038.        .......T......T.......GC.G......A.................   [298]
Pts039.        .......T......T........C.G......A.................   [298]
Pts040.        .......T......T........C.G......A.................   [298]
Pts041.        .......T......T........C.G......A.................   [298]
Pts043.        .......T......T........C.G......A.................   [298]
Pts044.        .......T......T........C.G......A.................   [298]
Pts045.        .......T......T........C.G......A.....G...........   [298]
Pts046.        .......T......T........C.G......A.............T...   [298]
Pts047.        .......T......T........C.G......A.............T...   [298]
```

```
Pts048.     .......T......T........C.G......A..................   [298]
Pts049.     .......T......T........C.G......A..................   [298]
Pts050.     .......T...............C.GC.....A..................   [298]
Pts051.     .......T...............C.GC.....A..................   [298]
Pts052.     .......T......T........C.G......A..................   [298]
Pts053.     .......T......T........C.G......A..................   [298]
Pts054.     .......T......T........C.G......A..................   [298]
Pts055.     .......T......T........C.G......A..................   [298]
Pts056.     .......T......T........C.G......A..................   [298]
Pts057.     .......T......T........C.G......A..................   [298]
Pts058.     .......T......T........C.G......A..................   [298]
Pts059.     .......T......T........C.G......A..................   [298]
Pts060.     .......T......T........C.G......A..................   [298]
Pts061.     .......T......T........C.G......A..................   [298]
Pts062.     .......T......T........C.G......A..................   [298]
Pts063.     .......T......T........C.G....T.A..................   [298]
Pts065.     .......T...............C.G......A..................   [298]
Pts066.     .......T......T........C.G......A..................   [298]
Pts067.     .......T......T........C.G......A..................   [298]
Pts068.     .......T......T........C.G.....TA..................   [298]
Pts069.     .......T......T........C.G......A..................   [298]
Pts070.     .......T......T........C.G......A..................   [298]
Pts071.     .......T......T........C.G....T.A..................   [298]
Pts072.     .......T......T........C.G......A..................   [298]
Pts073.     .......T......T........C.G......A..................   [298]
Pts074.     .......T......T........C.G......A..................   [298]
Pts075.     .......T......T........C.G......A..................   [298]
Pts076.     ..G....T......T........C.G......A..................   [298]
Pts077.     .......T......T........C.G......A..................   [298]
Pts078.     .......T......T........C.G......A..................   [298]
Pts079.     .......T......T........C.G......A..................   [298]
Pts080.     .......T......T........C.G......A.....G............   [298]
Pts081.     .......T......T........C.G......A..................   [298]
Pts082.     .......T......T........C.G......A..................   [298]
Pts083.     .......T......T........C.G......A..................   [298]
Pts084.     .......T......T........C.G......A..................   [298]
Pts085.     .......T......T........C.G......A..................   [298]
Pts086.     .......T......T........C.GC.....A..................   [298]
GOLCOMP01   .......T......T........C.G......A..................   [298]
Pts089.     .......T......T........C.G......A..................   [298]
Pts090.     .......T......T........C.G......A..................   [298]
Pts091.     .......T......T.......TC.G....T.A..................   [298]
Pts092.     .......T......T........C.G......A..................   [298]
Pts093.     .......T......T........C.G......A..................   [298]
Pts094.     .......T......T........C.G......A..................   [298]
Pts095.     .......T......T........C.G....T.A..................   [298]
Pts096.     .......T......T........C.G......A..................   [298]
Pts097.     .......T......T........C.G......A..................   [298]
Pts098.     .......T......T.......GC.G......A..................   [298]
Pts099.     .......T......T........C.G......A..................   [298]
Pts100.     .......T......T........C.G......A..................   [298]
Pts101.     .......T......T........C.G......A..................   [298]
Pts102.     .......T......T........C.G......A..................   [298]
Pts103.     .......T......T........C.G......A..................   [298]
```

```
Pts104.      .......T......T........C.G......A.................   [298]
Pts105.      .......T......T........C.G......A.................   [298]
Pts106.      .......T...............C.G......A.................   [298]
Pts107.      .......T......T........C.G......A.................   [298]
Pts108.      .......T......T........C.G......A.................   [298]
Pts109.      .......T......T........C.G......A.................   [298]
Pts110.      .......T......T........C.G......A.................   [298]
Pts111.      .......T......T........C.G......A.................   [298]
Pts112.      .......T......T........C.G......A.................   [298]
Pts113.      .......T......T........C.G......A.................   [298]
Pts114.      .......T......T........C.G......A.................   [298]
Pts115.      .......T......T........C.G......A.................   [298]
Pts116.      .......T......T........C.G......A.................   [298]
Pts117.      .......T......T........C.G......A.................   [298]
Pts118.      .......T......T........C.G......A.................   [298]
Pts119.      .......T......T........C.G......A.................   [298]
Pts120.      .......T......T........C.G......A.................   [298]
Pts121.      ...?...T...?.T.......?C.G......?.................   [298]
Pts122.      .......T......T........C.G......A.................   [298]
Pts123.      .......T......T........C.G......A.................   [298]

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

Human        TCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGGT   [345]
H45          TCAAATCCCTTCTTGTCCCCATGGATGACCCCCCTCAGATAG????   [345]
Ppaniscus    TCAAATCCATCCTCGCCCCCACGGATGCCCCCCCTCAGATAGGGGT   [344]
P_1Kp        TCAAATCCATCCTCGCCCCCACGGATGCCCCCCCTCAGATAGGAA?   [344]

CRNP2        TCAAATCCATCCTCGCCCCCACGGATGACCCCCCTCAGATAGGAGT   [345]
CRNP3        ...................-.......C.....?............   [344]
CRNP6        ...............T...-.......C..................   [343]
CRNP9        ..............................................   [345]
CRNP10       ...........................C..................   [345]
CRNP12       ..........T....T...-.......C..................   [343]
AKZN6        .................................?............   [345]
AKZN7        .................................?............   [345]
AKZN8        .................................?............   [345]
AKZN10       ...................-.......C.....?............   [344]
GONCOMP01    ...........................C..................   [345]
AKZN13       ...................-..........................   [344]
AKZN14       ...........................C..................   [345]
AKZN16       ...................-..........................   [344]
NGNY14       ..............................................   [345]
NGNY15       ..............................................   [345]
GGNP951      ...........................C..................   [345]
GGNP954      ...........................C.....?............   [345]
GGNP955      ....?..........T...........C..................   [345]
GGNP21       ..............................................   [345]
GGNP23       ...............T...........C................A.   [345]
GGNP27       ...................-.......C.....?............   [344]
BYM1         .............GC....-..........................   [344]
BYM3         .T...C.T..T....T...........C...............G..   [344]
BYM4         ..............................................   [345]
```

```
BYM5          ..............................................   [345]
MTCM2         ...........................C..................   [345]
MTCM4         ...........................C................A.   [345]
MTCM7         ...............T..............................   [345]
MTCM8         ..............................................   [345]
MTCM11        ...........................C................TC   [345]
MTCM12        ..............................................   [345]
MTCM15        ...................-.......C..................   [344]
MTCM16        ...................-..........................   [344]
MTCM21        ...............T...........?................A.   [345]
Mosse1        ...................-.............?..........A.   [344]
Mosse2        .................................?..........TC   [345]
Mosse5        .................................?............   [345]
Mosse8        .................................?............   [345]
Mosse12       ...........................C................A.   [345]
Mosse13       ..............................................   [345]
Mosse14       ...........................C................A.   [345]
Mosse15       ...............T...-..........................   [344]
Mosse21       ..............CG............................TC   [345]
Man1          ...............T..........-G..................   [343]
Man3          ...............T...........C..................   [344]
Man6          ...G.......TC..T..........TG................T.   [345]
Man7          ..............................................   [345]
Man8          ...............T...-.......C.....?............   [343]
Man10         ..............................................   [345]
Man11         ..............................................   [345]
Man12         ..............................................   [345]
Man15         ...................-..........................   [344]
Man16         ..............................................   [345]
DEFR1         ...............T...........C.........AG.......   [344]
DEFR2         ...............T...........C..................   [344]
DEFR3         ..........T..T.T...........C..................   [344]
DEFR6         .............T.T...........C...............G..   [345]
DEFR7         ..........T....T...........C................A.   [344]
DEFR8         ..........T....T...........C.....?............   [343]
DEFR9         ...............T..........TG..................   [344]
Campo1        ..........T....T...........C..................   [344]
Campo2        ...C......T....T...........C..................   [344]
Campo6        ..........T....T...........C..................   [344]
GONCOMP02     ..........T....T...........C..................   [344]
Campo9        ..........T..T.T...........C.....T.........GA.   [344]
Campo13       ..........T................C...............GA.   [344]
Campo14       ..........T....T...........C............TA.GA.   [344]
Campo15       ....T.....T................C.....?............   [344]
DBR2          .............T.T...-.......C.....?..........A.   [344]
DBR3          .............T.............C.....?.........GA.   [345]
DBR5          ...............T...........C.....?............   [345]
GONCOMP03     ...............T...........C..................   [345]
DBR7          ...........................C.....?............   [344]
DBR10         ..........T....T...........C.....?..........T.   [344]
DBR12         ..........T....T...........C................A.   [344]
DBR13         .............T.T...-.......C...............G..   [344]
ISEFR1        ...........................CT............?????   [345]
ISEFR3        ...........................CT....?.??.????????   [345]
```

```
Gimble         ...........................C...............G..   [344]
MORCOMP01      ...........................C...............G..   [344]
MORCOMP02      ...........................C...............G..   [344]
MORCOMP03      ...........................C...............G..   [344]
WL_3           ...........................C...............G..   [344]
MORCOMP04      ...........................C...............G..   [344]
George         ...........................C...............G..   [344]
Pts_43         ...............G...........C...............G..   [344]
MORCOMP05      ........G..................C...............G..   [344]
Gemini         ..........T....T...........C.................?   [344]
Ivindo         ..........T....T...........C.................?   [344]
Makokou        ..........T....T...........C..................   [344]
Mopia          ..........T................C.................?   [344]
Mouilla        ..........T....T..T........C..................   [344]
Ponia          .............T.............C.................?   [344]
Ptt_36         .............T.T...........C.................?   [346]
Ptt_37         .....C....T....T...........C..................   [344]
Berthe         .............T.T...........C.................?   [345]
Milla          .............T.T...........C...............G..   [346]
C1             ..........T....T...........C..................   [345]
Ptv_4          ............................T........?????????   [344]
Ptv_5          ......T.T......T...........C................A.   [344]
Ptv_15         .....CT.T......T...........C................A.   [345]
C2             ............................T...............A.   [345]
C3             .....C..C..................CT...............A.   [344]
GAGCOMP01      ............................T...............A.   [344]
GAGCOMP02      ......T.T......T...........C................A.   [345]
Ulsse8430      ?....CT.T......T...........C................A.   [345]
RICCInc58      .............T..............T...........G...A.   [344]
GAGCOMP03      .....C..C..................CT...............A.   [345]
13_p84339      .....CT.T......T...........C................A.   [345]
12_84303       .............T..............T...........G...A.   [344]
GAGCOMP04      .....CT.T......T...........C................A.   [344]
GAGCOMP05      .T...........T..............T...............A.   [343]
Ali            ......T.T......T...........C................A.   [344]
GAGCOMP6       .....CT.C......T...........C................A.   [344]
Darwin         .....CT.T......T...........C................A.   [344]
Dilly          ........................T...T...............A.   [344]
Hera           .....C..C......T...........CT...............A.   [345]
Loukoum        .............T..............T...............A.   [343]
Ondine         ..........T..T.T............T...........G...A.   [344]
Perla          .....??........T...-......?CT...............A.   [343]
Rousseau       .....CT.TC.....T...........G.............A..A.   [344]
Salome         .....C..C..................CT...............A.   [345]
Venus          .....C..C..................CT...............A.   [344]
Xeres          .....?..T......T...........CA...............A.   [345]
B3             .T...........T..............T...............A.   [344]
F18            .........................C..T...............A.   [344]
K22            ............................T...............A.   [344]
M19            .....C......................T...............A.   [344]
M18            .T...........T..............T............?..A.   [343]
Solo10         .T...........T..............T...............A.   [344]
Solo13         ...?..T.T......T...........C.....?..........A.   [344]
Wongo1         ......?.T......T...........C..............?.A.   [345]
```

```
Wongo7       ......T.T......T............T...............A.   [345]
Wongo5       .............T..............T.......?...?...?.   [344]
Pts001.      ..........T................G...............GA.   [344]
Pts002       ..........T................C.........A.....GA.   [344]
Pts003       ..........T................C.........A.....GA.   [344]
Pts005.      ..........T................C...............G..   [344]
Pts006.      ..........T................C...............G..   [344]
Pts007.      ..........T................C...............G..   [344]
Pts008.      ..........T................C...............G..   [344]
Pts009.      ..........T................C...............G..   [344]
Pts010.      ..........T................C...............G..   [344]
Pts011.      ..........T................C..........G....G..   [344]
Pts012.      ..........T................C..........G....G..   [344]
Pts013.      ..........T................C...............G..   [344]
Pts014.      ..........T................C.........AG....G..   [344]
Pts015.      ...........................C...............G..   [344]
Pts016.      ...........................C...............G..   [344]
Pts017.      ...........................C...............G..   [344]
Pts018.      ...........................C...............G..   [344]
Pts019.      ...........................C...............G..   [344]
Pts020.      ...........................C...............G..   [344]
Pts021.      ...........................C...............G..   [344]
Pts022.      ...........................C...............G..   [344]
Pts023.      .......T...................C...............G..   [344]
Pts024.      .......T...................C...............G..   [344]
Pts025.      ...........................C...............G..   [344]
Pts026.      ...........................C...............G..   [344]
Pts027.      ...........................C...............G..   [344]
Pts028.      ...........................C...............G..   [344]
Pts029.      ...........................T...............G..   [344]
Pts030.      ...........................C...............G..   [344]
Pts031.      ...........................C........?......G..   [344]
Pts032.      ...........................C...............G..   [344]
Pts033.      ...........................C...............G..   [344]
Pts034.      ...........................C...............G..   [344]
Pts035.      ...........................C...............G..   [344]
Pts036.      ...........................C...............G..   [344]
Pts037.      ...............T...........C...............G..   [344]
Pts038.      ...........................C...............G..   [344]
Pts039.      ...........................T...............G..   [344]
Pts040.      ...........................C...............G..   [344]
Pts041.      ...........................C...............G..   [344]
Pts043.      ...........................C...............G..   [344]
Pts044.      ...........................C...............G..   [344]
Pts045.      ...........................C...............G..   [344]
Pts046.      ...........................C...............G..   [344]
Pts047.      ..........T................C...............G..   [344]
Pts048.      ...........................C...............G..   [344]
Pts049.      ...........................C...............G..   [344]
Pts050.      ...........................T...............G..   [344]
Pts051.      ...........................C...............G..   [344]
Pts052.      ....................G......C....T..........G..   [344]
Pts053.      ...........................C....T..........G..   [344]
Pts054.      ...........................C....T..........G..   [344]
```

```
Pts055.      ........................C.............G..   [344]
Pts056.      ........................C.............G..   [344]
Pts057.      ........................C.............G..   [344]
Pts058.      ........................C.............G..   [344]
Pts059.      ........................C.............G..   [344]
Pts060.      ........................C.............G..   [344]
Pts061.      ........................C.............G..   [344]
Pts062.      ........................C.............G..   [344]
Pts063.      ........................C.............G..   [344]
Pts065.      ........................C.............G..   [344]
Pts066.      ........................C.............G..   [344]
Pts067.      ........................C.............G..   [344]
Pts068.      ........................C.............G..   [344]
Pts069.      ........................C...T.........G..   [344]
Pts070.      ........................C.............G..   [344]
Pts071.      ........................C.............G..   [344]
Pts072.      .........T..............C.............G..   [344]
Pts073.      ........................C.............G..   [344]
Pts074.      ........................C.............G..   [344]
Pts075.      ........................C.............G..   [344]
Pts076.      ........................C.............G..   [344]
Pts077.      ........................C.............G..   [344]
Pts078.      ........................C.............G..   [344]
Pts079.      ........................C.............G..   [344]
Pts080.      ........................C.............G..   [344]
Pts081.      ........................C.............G..   [344]
Pts082.      ........................C.............G..   [344]
Pts083.      ........................C.............GA.   [344]
Pts084.      ........................C.............G..   [344]
Pts085.      ........................C.............G..   [344]
Pts086.      ........................C.............G..   [344]
GOLCOMP01    ........................C.............G..   [344]
Pts089.      ........................C.............G..   [344]
Pts090.      ........................C.............G..   [344]
Pts091.      ........................C.............G..   [344]
Pts092.      ........................C................   [344]
Pts093.      ........................C.............G..   [344]
Pts094.      ........................C.............G..   [344]
Pts095.      ........................C.............G..   [344]
Pts096.      ........................C.............G..   [344]
Pts097.      ........................C.............G..   [344]
Pts098.      ........................C.............G..   [344]
Pts099.      ........................C.............G..   [344]
Pts100.      ........................C.............G..   [344]
Pts101.      ........................C.............G..   [344]
Pts102.      ........................C.............G..   [344]
Pts103.      ........................C.............G..   [344]
Pts104.      ........................C.............G..   [344]
Pts105.      .........T..............C.............G..   [344]
Pts106.      ........................C.............G..   [344]
Pts107.      ........................C.............G..   [344]
Pts108.      ........................C.............G..   [344]
Pts109.      ........................C.............G..   [344]
Pts110.      ........................C.............G..   [344]
```

```
Pts111.       ...........................C...............G..   [344]
Pts112.       ...........................C...............G..   [344]
Pts113.       ...........................C...............G..   [344]
Pts114.       ...........................C...............G..   [344]
Pts115.       ...........................C...............G..   [344]
Pts116.       ...........................C...............G..   [344]
Pts117.       ...........................C...............G..   [344]
Pts118.       ...........................T...............G..   [344]
Pts119.       ...........................T...............G..   [344]
Pts120.       ...........................C...............G..   [344]
Pts121.       ...........................C...............G..   [344]
Pts122.       ...........................C...............G..   [344]
Pts123.       ...........................C...............G..   [344]
```

Composite Haplotypes

Sample Name	Composite Sequence Composition	Geographic Region(s)	Author
GONCOMP01	AKZN12, NNFR951, GGNP8, GGNP11	Nigeria, Cameroon Highlands	Gonder *et al*, 1997, Gonder, this study
GONCOMP02	Campo7, Campo10, Campo12	Southern Cameroon, Campo Forest Reserve	Gonder, present study
GONCOMP03	DBR6, DBR8, DBR16	Southern Cameroon, Dja Biosphere Reserve	Gonder, present study
MORCOMP01	Evered, Cal, TT_3	Eastern Africa, Gombe	Morin *et al*, 1994
MORCOMP02	Frodo, Kidveu	Eastern Africa, Gombe	Morin *et al*, 1994
MORCOMP03	Prof, JF_1b, Pts_49	Eastern Africa, Gombe	Morin *et al*, 1994
MORCOMP04	BE_1, AL_2	Eastern Africa, Gombe	Morin *et al*, 1994
MORCOMP05	Ptt_43, Ntebe	Western Equatorial Africa	Morin *et al*, 1994
GAGCOMP01	12_84303, Afua	Upper Guinea, Ghana	Gagneux, 1998, Gonder, present study
GAGCOMP02	Gitane842, Gala, Tobar	Upper Guinea	Gagneux, 1998
GAGCOMP03	K3, 7_84312	Upper Guinea	Gagneux, 1998
GAGCOMP04	Solo_11, Solo_84331	Upper Guinea	Gagneux, 1998
GAGCOMP05	B12, Belle	Upper Guinea	Gagneux, 1998
GOLCOMP01	Pts087, Pts088	Eastern Africa	Goldberg, 1996

Note:

1. The 254 ingroup taxa base pairs are referenced against CRNP2. HV-I DNA sequences from other studies can be found in Genbank under the following accession numbers: L35381-L35443, U77186-U77293, AFO59042-AFO59052 and AF137481-AF137412.

APPENDIX 3

Identification of Samples shown in Phylogenetic Trees

Taxon Number	Sample Name			
	Parsimony	Maximum Likelihood	Neighbor-Joining Jackknife	Transversion “uncorrected p”
1.	ISEFR3	ISEFR1	ISEFR3	13p84339
2.	ISEFR1	ISEFR3	ISEFR1	Ali
3.	Hera	Man15	Venus	Darwin
4.	Venus	Mosse13	Salome	C3
5.	Salome	Mosse21	GAGCOMP03	Ptv15
6.	GAGCOMP03	Mosse2	Hera	Wongo7
7.	C3	Mosse1	C3	Solo13
8.	Xeres	Man7	Xeres	Rousseau
9.	GAGCOMP06	Man10	Rousseau	GAGCOMP02
10.	Ali	BYM5	Darwin	Xeres
11.	13p84339	MTCM16	GAGCOM04	GAGCOMP3
12.	Rousseau	BYM4	13p84339	GAGCOMP04
13.	Darwin	AKZN13	Ulsse8430	Ulsse8430
14.	GAGCOMP04	CRNP9	GAGCOMP06	GAGCOMP06
15.	Ulsee8430	AKZN6	Ptv15	Venus
16.	GAGCOMP02	Man16	Wongo1	Salome
17.	Ptv15	Man11	Wongo7	Hera
18.	Wongo7	Man12	Solo13	Ptv5
19.	Wongo1	Mosse15	Ali	Mosse14
20.	Solo13	MTCM12	GAGCOMP02	ISEFR3
21.	Ptv5	BYM1	Ptv 5	MTCM11
22.	Wongo5	NGNY15	Wongo5	ISEFR1
23.	Ondine	NGNY14	12 84303	MTCM4
24.	12 84303	Mosse5	Ondine	Mosse12
25.	RICCInc58	GGNP951	RICCInc58	GGNP23
26.	K22	GGNP955	F18	DBR7
27.	Loukoum	AKZN14	Dilly	MTCM2
28.	Solo10	Mosse8	Solo 10	GONCOMP01
29.	M18	AKZN16	K22	GGNP27
30.	B3	GGNP21	Loukoum	MTCM15
31.	GAGCOMP05	AKZN8	M18	GGNP954
32.	M19	AKZN7	B3	AKZN10
33.	F18	MTCM8	GAGCOMP05	GGNP951

Taxon Number	Sample Name			
	Parsimony	Maximum Likelihood	Neighbor-Joining Jackknife	Transversion "uncorrected p"
34.	Perla	CRNP2	Perla	GGNP955
35.	Dilly	CRNP10	M19	CRNP10
36.	GAGCOMP01	CRNP3	GAGCOMP01	CRNP3
37.	C2	DBR7	C2	Perla
38.	Ptv4	MTCM15	Ptv4	MTCM21
39.	MTCM7	MTCM2	CRNP2	Man7
40.	Mosse14	GGNP954	MTCM8	AKZN14
41.	Mosse12	GONCOMP01	CRNP9	Man11
42.	MTCM11	GGNP27	AKZN6	MTCM7
43.	MTCM4	AKZN10	MTCM16	GGNP21
44.	MTCM21	Mosse12	BYM5	Mosse15
45.	Man6	Mosse14	Man10	AKZN8
46.	GGNP23	MTCM7	Man11	BYM5
47.	DBR7	MTCM11	Man16	Man10
48.	MTCM15	MTCM4	BYM1	Man16
49.	MTCM2	MTCM21	MTCM12	AKZN7
50.	GGNP27	Man6	Mosse15	CRNP9
51.	GGNP954	GGNP23	Man7	MTCM12
52.	GONCOMP01	GAGCOMP03	Man12	MTCM8
53.	AKZN10	Venus	Mosse5	BYM4
54.	CRNP10	Salome	BYM4	Mosse13
55.	CRNP3	Hera	AKZN7	Man12
56.	Man15	C3	AKZN8	Mosse2
57.	Mosse13	Xeres	GGNP21	Mosse21
58.	Man10	13 p84339	AKZN16	BYM1
59.	MTCM16	Ulsse8430	Mosse8	AKZN13
60.	BYM5	Rousseau	Mosse13	MTCM16
61.	Man16	Darwin	Man15	Man15
62.	Man11	GAGCOMP04	AKZN14	Mosse8
63.	BYM4	GAGCOMP06	GGNP951	Mosse5
64.	AKZN13	Ptv 15	GGNP955	Mosse1
65.	AKZN6	Wongo7	Mosse1	AKZN16
66.	CRNP9	Wongo1	Mosse2	AKZN6
67.	Mosse5	Solo13	Mosse21	CRNP2
68.	Mosse15	GAGCOMP02	NGNY14	B3
69.	MTCM12	Ali	NGNY15	NGNY15
70.	Man12	Ptv5	AKZN13	NGNY14
71.	Man7	Wongo5	CRNP3	GAGCOMP01
72.	BYM1	12 84303	CRNP10	RICCInc58
73.	NGNY15	Ondine	AKZN10	F18
74.	NGNY14	RICCInc58	GONCOMP01	GAGCOMP05

Taxon Number	Sample Name			
	Parsimony	Maximum Likelihood	Neighbor-Joining Jackknife	Transversion "uncorrected p"
75.	Mosse21	Loukoum	GGNP27	Ondine
76.	Mosse2	Solo 10	GGNP954	K22
77.	Mosse1	M18	MTCM2	Loukoum
78.	GGNP955	B3	MTCM15	M19
79.	GGNP951	GAGCOMP05	DBR7	Dilly
80.	AKZN14	K22	GGNP23	Solo10
81.	Mosse8	M19	MTCM21	C2
82.	AKZN16	Perla	Man6	Man6
83.	GGNP21	F18	MTCM4	Wongo5
84.	AKZN8	Dilly	MTCM11	Ptv4
85.	AKZN7	GAGCOMP01	Mosse12	WL3
86.	MTCM8	C2	Mosse14	Pts031
87.	CRNP2	Ptv 4	MTCM7	MORCOMP04
88.	Campo9	Pts070	Milla	Pts041
89.	Milla	Pts069	DBR3	MORCOMP02
90.	DBR13	Campo13	GONCOMP03	Pts032
91.	DEFR6	Campo15	DBR5	Pts043
92.	DBR3	DEFR3	DBR2	Pts044
93.	DBR2	GONCOMP02	Berthe	Pts037
94.	GONCOMP03	Campo14	DBR13	Pts038
95.	DBR5	Campo1	DEFR6	Pts036
96.	Berthe	CRNP12	Ptt 36	Pts039
97.	Ptt36	DEFR9	Campo15	Pts033
98.	Pts018	DEFR2	Campo13	Pts114
99.	Pts017	DEFR1	Ponia	Pts034
100.	Pts010	CRNP6	Campo14	Pts035
101.	Pts009	C1	DEFR3	Pts028
102.	Pts008	DBR10	Campo1	Pts029
103.	Pts007	Campo6	GONCOMP02	Pts030
104.	Pts006	Campo2	CRNP12	Pts091
105.	Pts005	Ptt 37	DEFR9	BYM3
106.	MORCOMP01	Ponia	DEFR2	Pts045
107.	Pts016	Mopia	DEFR1	DBR13
108.	Pts015	Makokou	CRNP6	DEFR6
109.	Pts021	Ivindo	Campo6	DEFR9
110.	Pts020	Man8	Campo2	Man1
111.	Pts019	Man3	Ptt 37	DBR12
112.	Pts027	Man1	Mopia	CRNP12
113.	Pts026	DEFR8	DBR10	GONCOMP02
114.	Pts024	DBR12	Man8	Campo13
115.	Pts023	DEFR7	Man3	Man3

Taxon Number	Sample Name			
	Parsimony	Maximum Likelihood	Neighbor-Joining Jackknife	Transversion “uncorrected p”
116.	Pts022	Mouilla	Man1	Campo2
117.	Pts025	Gemini	C1	DEFR7
118.	MORCOMP03	MORCOMP05	Makokou	DEFR1
119.	Pts121	Pts086	Ivindo	Campo14
120.	Pts120	Pts045	DBR12	C1
121.	Pts119	Pts047	DEFR7	DBR10
122.	Pts118	Pts046	Mouilla	Campo15
123.	Pts117	Pts044	DEFR8	Campo6
124.	Pts116	Pts043	Gemini	Ptt 37
125.	Pts110	Pts041	MORCOMP05	DEFR8
126.	Pts104	MORCOMP04	Campo9	Man8
127.	Pts103	BYM3	Pts009	Makokou
128.	Pts101	Pts035	Pts008	Mouilla
129.	Pts102	Pts038	Pts010	DEFR3
130.	Pts100	Pts114	Pts007	DEFR2
131.	Pts099	Pts039	Pts006	Mopia
132.	Pts098	Pts037	Pts005	Ivindo
133.	Pts097	Pts036	Pts017	Ponia
134.	Pts096	Pts033	Pts018	Campo1
135.	Pts094	Pts034	MORCOMP01	CRNP6
136.	Pts093	Pts030	Pts016	Gemini
137.	Pts092	Pts029	Pts015	MORCOMP05
138.	Pts095	Pts028	Pts021	Pts121
139.	Pts091	Pts031	Pts020	Pts043
140.	Pts090	WL 3	Pts019	Pts001
141.	Pts089	Pts032	Pts027	Pts076
142.	GOLCOMP01	MORCOMP02	Pts026	MORCOMP03
143.	Pts086	Campo9	Pts024	Pts025
144.	Pts084	Milla	Pts023	Pts012
145.	Pts083	DBR13	Pts022	Pts003
146.	Pts082	DEFR6	Pts025	Pts002
147.	Pts081	DBR3	MORCOMP03	Pts068
148.	Pts080	DBR2	Pts045	Pts022
149.	Pts079	GONCOMP03	Pts086	Pts017
150.	Pts078	DBR5	Pts070	Pts083
151.	Pts077	Bertha	Pts069	Pts086
152.	Pts076	Ptt 36	Pts068	Pts069
153.	Pts073	Pts021	Pts003	Pts070
154.	Pts075	Pts020	Pts002	Pts109
155.	Pts072	Pts019	Pts001	Pts079
156.	Pts071	Pts024	BYM3	Pts084

Taxon Number	Sample Name			
	Parsimony	Maximum Likelihood	Neighbor-Joining Jackknife	Transversion “uncorrected p”
157.	Pts070	Pts023	Pts044	Pts073
158.	Pts069	Pts027	Pts043	Pts089
159.	Pts068	Pts026	Pts041	Pts092
160.	Pts067	Pts022	MORCOMP04	Pts110
161.	Pts066	Pts025	Pts039	Pts016
162.	Pts051	MORCOMP03	Pts114	Pts106
163.	Pts050	Pts016	Pts038	Pts023
164.	Pts048	Pts015	Pts037	Pts058
165.	Pts074	Pts010	Pts036	Pts078
166.	Pts47	Pts008	Pts035	Pts108
167.	Pts046	Pts007	Pts028	Pts111
168.	Pts045	Pts009	Pts031	Pts118
169.	Pts115	Pts006	Pts030	Pts120
170.	Pts113	Pts005	Pts029	Pts105
171.	Pts111	Pts017	WL 3	Pts123
172.	Pts109	Pts018	Pts034	Pts119
173.	Pts108	MORCOMP01	Pts033	Pts101
174.	Pts106	George	Pts032	Pts05
175.	Pts105	Pts078	MORCOMP02	Pts019
176.	Pts107	Pts076	Pts117	Pts107
177.	Pts049	Pts077	Pts119	Pts059
178.	Pts112	Pts073	Pts118	Pts122
179.	Pts085	Pts075	Pts120	Pts066
180.	Pts040	Pts084	Pts066	Pts053
181.	Pts123	Pts083	Pts077	Pts049
182.	Pts104	Pts081	Pts073	Pts052
183.	Pts013	Pts082	Pts072	Pts080
184.	Pts012	Pts080	Pts071	Pts026
185.	Pts011	Pts079	Pts075	Pts010
186.	Pts003	Pts074	Pts084	Pts006
187.	Pts002	Pts072	Pts083	Pts008
188.	Pts001	Pts071	Pts081	Pts011
189.	BYM3	Pts119	Pts082	Pts098
190.	Ponia	Pts118	Pts080	Pts099
191.	Mopia	Pts117	Ps079	Pts097
192.	Campo6	Pts113	Pts074	Campo9
193.	Campo2	Pts115	Pts047	Pts115
194.	Ptt37	Pts111	Pts046	Pts077
195.	DBR10	Pts105	Pts078	Pts024
196.	Man8	Pts109	Pts076	Pts116
197.	Man3	Pts108	George	Pts112

Taxon Number	Sample Name			
	Parsimony	Maximum Likelihood	Neighbor-Joining Jackknife	Transversion "uncorrected p"
198.	Man1	Pts106	Pts109	Pts102
199.	Campo15	Pts107	Pts105	Pts096
200.	Campo13	Pts049	Pts106	Pts075
201.	Campo14	Pts112	Pts108	Pts071
202.	GONCOMP02	Pts085	Pts107	Pts050
203.	Campo1	Pts040	Pts49	Pts040
204.	DEFR3	Pts123	Pts112	Pts021
205.	CRNP12	Pts068	Pts085	Pts020
206.	DEFR9	Pts002	Pts113	Pts063
207.	DEFR2	Pts003	Pts115	Pts090
208.	DEFR1	Pts001	Pts111	GOLCOMP01
209.	CRNP6	Pts110	Pts040	Pts100
210.	C1	Pts116	Pts123	Pts095
211.	Makokou	Pts099	Pts013	Pts093
212.	Ivindo	Pts098	Pts014	Pts074
213.	DEFR8	Pts097	Pts012	Pts072
214.	DBR12	Pts103	Pts011	Pts048
215.	DEFR7	Pts120	Pts116	Pts060
216.	Mouilla	Pts066	Pts110	Pts056
217.	Gemini	Pts101	Pts090	Pts104
218.	MORCOMP05	Pts102	Pts089	Pts103
219.	Pts056	Pts100	GOLCOMP01	Pts085
220.	Pts122	Pts090	Pts099	Pts014
221.	Pts063	Pts089	Pts098	Pts067
222.	Pts062	GOLCOMP01	Pts097	Pts013
223.	Pts061	Pts094	Pts102	Pts094
224.	Pts060	Pts096	Pts100	Pts047
225.	Pts059	Pts092	Pts101	Pts046
226.	Pts058	Pts093	Pts093	GONCOMP03
227.	Pts057	Pts095	Pts096	DBR5
228.	Pts055	Pts091	Pts094	DBR2
229.	Pst054	Pts048	Pts092	DBR3
230.	Pts053	Pts014	Pts095	Berthe
231.	Pts052	Pts013	Pts091	Pts065
232.	Pts065	Pts012	Pts103	Pts061
233.	Pts043	Pts011	Pts048	Pts054
234.	George	Pts104	Pts104	Pts062
235.	Pts044	Pts067	Pts067	Pts117
236.	Pts043	Pts056	Pts051	Pts057
237.	Pts041	Pts055	Pts050	Pts015
238.	MORCOMP04	Pts054	Pts056	Pts081

Taxon Number	Sample Name			
	Parsimony	Maximum Likelihood	Neighbor-Joining Jackknife	Transversion "uncorrected p"
239.	Pts114	Pts053	Pts055	Pts082
240.	Pts039	Pts052	Pts060	Pts027
241.	Pts038	Pts058	Pts059	Pts113
242.	Pts037	Pts057	Pts058	Pts009
243.	Pts036	Pts059	Pts057	Pts007
244.	Pts035	Pts060	Pts063	Pts055
245.	Pts034	Pts051	Pts061	Pts018
246.	Pts031	Pts050	Pts062	Ptt 36
247.	Pts030	Pts062	Pts054	Pts005
248.	Pts029	Pts112	Pts053	Milla
249.	Pts028	Pts061	Pts052	George
250.	WL 3	Pts063	Pts065	MORCOMP01
251.	Pts033	Pts065	Pts043	Gimble
252.	Pts032	Pts043	Pts122	P 1kp
253.	MORCOMP02	Pts121	Pts121	Ppaniscus
254.	Gimble	Gimble	Gimble	H45
255.	P 1kp	P 1kp	P 1kp	Human
256.	Ppaniscus	Ppaniscus	Ppaniscus	
257.	H45	H45	H45	
258.	Human	Human	Human	

APPENDIX 4

STR Genotypes of Chimpanzees in Nigeria and Cameroon

Locus	APOA2		D16S265		HUMPLA2A		D4S1652		D7S1809	
Nigeria										
Akoh Zanto										
AKZN2	130	130			92	92			202	210
AKZN3	128	128	80	106	86	86	130	130		
AKZN6	132	132	106	108	86	86	138	138	210	210
AKZN7	128	132	106	108	86	86	126	162	202	210
AKZN8	132	132	90	106	86	86	114	114	214	214
AKZN10	132	134	100	106			114	150		
AKZN11	132	134	106	108			150	154	218	222
AKZN12	130	132	106	108	80	80	122	122		
AKZN14	130	134	106	108			122	138		
AKZN15	130	134	104	110	86	86	122	122		
AKZN16			114	114	86	86	122	142	214	214
Cross River										
CRNP2	124	132					122	138	210	210
CRNP3	130	132			95	95	122	122	210	210
CRNP4	130	132			86	86	122	142	210	210
CRNP5	130	132			86	86	126	138	210	210
CRNP6	130	132					126	142		
CRNP8	126	132					130	130	210	210
CRNP9	132	134					138	138		
CRNP10	132	132					126	138		
CRNP11	130	132					126	134		
CRNP12	132	134					126	142	206	210
Gashaka Gumti										
GGNP3	130	132					118	118		
GGNP8	130	130			83	83			210	210
GGNP11	136	142	82	106	86	89	146	162	218	226
GGNP21	130	132	106	108	86	89	118	138	202	210
GGNP23										
GGNP27	130	132	82	108	77	86	154	158	202	210
GGNP28	134	136			83	86	126	138	226	230
GGNP32	134	136								
GGNP34	128	130	106	108	83	83	122	134		
GGNP38	130	132	80	106			126	134		

Locus	APOA2		D16S265		HUMPLA2A		D4S1652		D7S1809	
Ise										
ISEFR1	130	134					118	118	218	218
ISEFR3	130	132	78	106	80	89	138	138	206	206
ISEFR4	134	140	106	110			142	142	202	202
ISEFR5	130	132	80	108			122	134	206	206
Ngel Nyaki										
NGNY5	134	136	108	114	80	86	130	138		
NGNY9	130	132							202	202
NGNY10	136	138					154	154		
NGNY14	136	140	106	108			142	150		
NGNY15	130	132			86	86	122	146		
Skins										
Abong-sk	150	162	84	84	86	86			214	214
Ala-sk	120	154	96	102	89	89			214	218
Bomadi-sk										
Owo-sk										
Ise-sk	132	132					118	122		
Banyang Mbo										
BYM1	132	132	84	84	80	86	126	142	218	222
BYM3	132	132			80	88			210	210
BYM4	130	132	92	92	96	96	126	142		
BYM5	130	132			93	96	122	122		
BYM6	132	132			86	86				
Campo										
Campo1	132	134	114	114			126	138		
Campo2	130	130					122	122		
Campo3	132	140			86	86				
Campo5	126	126			83	83	118	146		
Campo6	132	136	108	110			122	150	230	242
Campo7										
Campo9	126	126					118	118		
Campo10	132	134			83	86			238	238
Campo12	132	134	94	102	86	86	130	130	238	242
Campo13	132	134			98	98	130	130		
Campo14	130	132					142	148	218	230
Campo15	132	134					134	134		
Dja										
DBR1							122	146		
DBR2	132	140	106	108						
DBR3	136	140								
DBR4	132	134	106	108	89	89				

Locus	APOA2		D16S265		HUMPLA2A		D4S1652		D7S1809	
DBR5	134	136			86	86				
DBR6	130	132	80	106	98	98			242	242
DBR7	132	134	106	110						
DBR8										
DBR10	132	140	106	108	86	86			210	242
DBR12			78	108						
DBR13	132	140					142	142		
DBR14										
DBR16	132	134	106	108			146	146	210	222
Douala-Edea										
DEFR1	132	134	80	106	86	89	122	122	226	226
DEFR2	132	134	106	108	89	89	118	122	226	226
DEFR3			106	108	86	86			210	210
DEFR4	130	132	80	108	86	86			210	210
DEFR5	132	140								
DEFR6					86	86				
DEFR7	128	132	80	108	86	86	138	138		
DEFR8	132	140			69	86	150	158		
DEFR9	130	134	80	108	86	86			238	242
DEFR10	132	132	80	108	86	86			210	210
DEFR11	134	136			98	98				
Man'bra										
Man1	120	120	80	108	77	77	122	122		
Man2					95	95				
Man3	132	132					122	150	218	242
Man4	126	126							214	214
Man6			90	108	86	86				
Man7	130	130	106	108						
Man8	130	132	80	106	86	95				
Man10			106	108			118	122		
Man11	128	132	106	108			118	122		
Man12			82	82						
Man13					77	77			206	210
Man14	130	132			86	86	118	122	218	218
Man15	126	132	80	108	86	86				
Man16	128	134	84	106						
Mosse										
Mosse1	124	132	82	108	86	86				
Mosse2										
Mosse5	130	132			86	86				
Mosse8			84	108						
Mosse11			80	104	86	86			214	214
Mosse12			80	108	86	86			210	210

Locus	APOA2		D16S265		HUMPLA2A		D4S1652		D7S1809	
Mosse13										
Mosse 14	130	134							214	218
Mosse15	128	136	80	108	65	65				
Mosse16	134	138	112	112	86	86				
Mosse20			80	108	86	92			214	218
Mosse21	130	138	80	106	80	89			206	230
Mount Cameroon										
MTCM2			80	108	86	92			234	238
MTCM4	128	130	80	106	86	95				
MTCM7										
MTCM8	136	148	80	110	86	89			206	210
MTCM10	126	126	80	108	86	98			202	202
MTCM11			80	108	86	86				
MTCM12			80	108	86	95				
MTCM15			106	108	80	86				
MTCM16	140	140	110	114	95	95				
MTCM20										
MTCM21	134	136	108	110	88	92			210	210
Pandrillus										
Nigeria										
Pan1										
Pan2										
Abu	134	142								
Daisy										
Itambu										
Jaki							154	154		
Kingsley										
Mickey							134	134		
Murphy										
Paablo	132	132					126	154		
Pansy	134	138								
Shirley	134	138								
Utangah	130	138								
Cameroon										
Akwaya										
Carlos	134	134					154	158		
Dja										
Ewake										
Jack							130	130		
Louisa										

Locus	APOA2		D16S265		HUMPLA2A		D4S1652		D7S1809	
Mac										
Margaret										
Maya							158	158		
Mokolo	130	132					126	158		
Po	130	132			86	86				
Poldina	130	134			95	101	122	126	214	218
Seko	124	132			88	92	118	122		
Somba	130	130			80	80				
Suzi	130	132			98	98	114	114		
Tobi	138	144			86	86	122	122		

Locus	D9S303		D11S1984		D13S317		D16S539		D20S470	
Nigeria										
Akoh Zanto										
AKZN2					175	191	152	164	279	284
AKZN3	150	162			183	187	152	152	247	263
AKZN6	150	154	188	188	175	183	160	164	227	287
AKZN7	162	162	180	180	183	187	160	164	303	303
AKZN8	166	166	184	184	191	191	152	152		
AKZN10			164	164			160	164	243	259
AKZN11	186	186			191	191	152	152	259	263
AKZN12			164	164			156	164	267	275
AKZN14	178	178	104	104			152	160		
AKZN15	190	190					152	164		
AKZN16			144	188			152	156	263	271
Cross River										
CRNP2	194	194	192	192	179	179	152	164	203	227
CRNP3					175	175	160	164	255	263
CRNP4							152	152	203	203
CRNP5							152	164	203	203
CRNP6	170	170					152	152		
CRNP8	162	162					148	164		
CRNP9	190	194			191	195	152	152	279	279
CRNP10							152	164	263	263
CRNP11	154	154			155	155	152	152	203	203
CRNP12							152	152		
Gashaka Gumti										
GGNP3	174	174					152	156	247	267
GGNP8	170	182	128	168	199	199	152	152	295	295

Locus	D9S303		D11S1984		D13S317		D16S539		D20S470	
GGNP11	174	190	168	168	187	187	152	156	287	295
GGNP21					223	227	152	164	251	267
GGNP23										
GGNP27					195	215	152	164	215	215
GGNP28					215	227	152	152		
GGNP32	170	178	168	168	195	215	152	164		
GGNP34			164	188	191	191	152	164		
GGNP38			124	188	191	195	152	152		
Ise										
ISEFR1	170	170	164	164	179	183	152	152	263	267
ISEFR3	174	174	192	192			152	152		
ISEFR4	174	174	188	188			152	156	267	283
ISEFR5	170	170					152	152		
Ngel Nyaki										
NGNY5										
NGNY9	186	190					152	160		
NGNY10										
NGNY14	166	170					152	152	259	263
NGNY15							152	152	243	247
Skins										
Abong-sk							156	168	287	287
Ala-sk									271	275
Bomadi-sk										
Owo-sk										
Ise-sk										
Banyang Mbo										
BYM1			176	176	191	195	152	156	259	263
BYM3							152	152	263	263
BYM4					155	155	160	164	247	283
BYM5							152	152	203	203
BYM6										
Campo										
Campo1							164	164		
Campo2	166	170					156	160	215	215
Campo3							152	152	167	167
Campo5			176	176	179	183	152	152	247	247
Campo6	146	170	180	180	183	183	152	152	267	267
Campo7										
Campo9	166	166								
Campo10							152	164		
Campo12	170	186	184	196	191	191	152	152	279	279
Campo13					187	191	152	164	267	283

Locus	D9S303		D11S1984		D13S317		D16S539		D20S470	
Campo14	170	170			179	179	164	164		
Campo15	170	174	184	184	179	183	152	164	251	263
Dja										
DBR1	170	170			179	179	152	152	283	283
DBR2	182	186					152	152	259	283
DBR3							152	164	275	279
DBR4	162	178	188	188	187	187	152	156	275	279
DBR5			184	184			152	152	259	263
DBR6	170	170	188	188			152	164	247	247
DBR7	170	170	180	180			152	152	255	259
DBR8										
DBR10	158	174	184	184	183	191	148	164	203	203
DBR12							152	156		
DBR13							152	152	203	215
DBR14										
DBR16	150	150					152	152	255	259
Douala-Edea										
DEFR1	146	146	168	168	187	191	152	156		
DEFR2	178	182	176	184	191	191	152	164	283	283
DEFR3	142	146	184	184			152	164	207	259
DEFR4	146	182	180	180			152	164	267	283
DEFR5							152	152	259	263
DEFR6							152	164	263	283
DEFR7			192	196	251	255			203	255
DEFR8	154	170			199	203	152	152	287	287
DEFR9	174	178			175	175	152	152		
DEFR10	162	162					152	164	263	263
DEFR11										
Man'bra										
Man1	170	170			247	255	152	152	283	287
Man2	158	162					156	156		
Man3					167	167	148	164	203	215
Man4							148	148		
Man6	166	170					152	164	263	283
Man7					183	187	152	164	263	283
Man8	154	162	180	180	191	191	152	164	247	283
Man10							152	164	267	275
Man11					179	179	152	164	283	291
Man12					175	175	152	152		
Man13					175	195	152	164	275	279
Man14							152	156	203	211
Man15							156	164	203	283
Man16	158	162					152	164	283	287

Locus	D9S303		D11S1984		D13S317		D16S539		D20S470	
Mosse										
Mosse1	174	182	168	168	171	191	156	160	215	283
Mosse2										
Mosse5			168	168			152	164	263	287
Mosse8					167	167	152	164		
Mosse11					247	247	152	164	263	279
Mosse12							144	160	259	259
Mosse13										
Mosse 14							156	156	227	227
Mosse15							152	152	279	287
Mosse16							152	152		
Mosse20							152	152		
Mosse21					179	199				
Mount Cameroon										
MTCM2					187	191	152	164	259	283
MTCM4					171	171	156	164	283	287
MTCM7					167	167	152	160		
MTCM8							152	164	263	283
MTCM10							152	164		
MTCM11					187	187	152	152		
MTCM12	162	162			175	187	152	156	259	263
MTCM15					183	191	156	156	207	279
MTCM16					175	187	152	152	203	215
MTCM20							148	152		
MTCM21							152	164	283	283
Pandrillus										
Nigeria										
Pan1										
Pan2										
Abu										
Daisy										
Itambu										
Jaki							156	160		
Kingsley										
Mickey							152	164		
Murphy										
Paablo										
Pansy										
Shirley	190	194	168	168			152	164	267	275
Utangah										

Locus	D9S303		D11S1984		D13S317		D16S539		D20S470	
Cameroon										
Akwaya										
Carlos										
Dja										
Ewake										
Jack					199	199				
Louisa										
Mac										
Margaret										
Maya										
Mokolo										
Po										
Poldina	174	190	180	200					267	283
Seko	170	170	168	192	187	187			271	283
Somba	174	178	168	188					287	287
Suzi	146	166	180	184	191	191			275	291
Tobi	142	162	168	180	183	183			263	291

APPENDIX 5

STR Summary Statistics

Sample location abbreviations and locations listed in Table 2.1 and Figure 2.2.

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
Western Nigeria				GGNP				NGNY				AKZN			
APOA2				APOA2				APOA2				APOA2			
120	1	0.08	0.08	128	1	0.06	0.06	130	2	0.2	0.13	128	3	0.15	0.08
130	3	0.25	0.13	130	7	0.39	0.19	132	2	0.20	0.13	130	5	0.25	0.10
132	4	0.33	0.14	132	4	0.22	0.10	134	1	0.10	0.10	132	8	0.40	0.11
134	2	0.17	0.11	134	2	0.11	0.08	136	3	0.30	0.15	134	4	0.20	0.09
140	1	0.08	0.08	136	3	0.17	0.09	138	1	0.10	0.10				
154	1	0.08	0.08	142	1	0.06	0.06	140	1	0.10	0.10				
Alleles			6				6				6				4
Genes			12				18				10				20
Heterozygosity			0.85				0.80				0.88				0.75
Exact HWE p =			0.39				0.05				0.63				0.42
D4S1652				D4S1652				D4S1652				D4S1652			
118	3	0.30	0.15	118	3	0.21	0.11	122	1	0.13	0.13	114	3	0.15	0.08
122	2	0.20	0.13	122	1	0.07	0.07	130	1	0.13	0.13	122	6	0.30	0.10
134	1	0.10	0.10	126	2	0.14	0.10	138	1	0.13	0.13	126	1	0.05	0.05
138	2	0.20	0.13	134	2	0.14	0.10	142	1	0.13	0.13	130	2	0.10	0.07
142	2	0.20	0.13	138	2	0.14	0.10	146	1	0.13	0.13	138	3	0.15	0.08
				146	1	0.07	0.07	150	1	0.13	0.13	142	1	0.05	0.05
				154	1	0.07	0.07	154	2	0.25	0.16	150	2	0.10	0.07
				158	1	0.07	0.07					154	1	0.05	0.05
				162	1	0.07	0.07					162	1	0.05	0.05
Alleles			5				9				7				9
Genes			10				14				8				20
Heterozygosity			0.87				0.93				0.96				0.89
Exact HWE p =			0.06				0.52				0.17				0.01

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
Western Nigeria				GGNP				NGNY				AKZN			
D7S1809				D7S1809				D7S1809				D7S1809			
202	2	0.20	0.13	202	2	0.20	0.13	202	2	1.00	0.00	202		0.17	0.11
206	4	0.40	0.16	210	4	0.40	0.16					210		0.33	0.14
214	3	0.10	0.10	218	1	0.10	0.10					214		0.33	0.14
218	1	0.30	0.15	226	2	0.20	0.13					218		0.08	0.08
				230	1	0.10	0.10					222		0.08	0.08
Alleles			4				5								5
Genes			10				10								12
Heterozygosity			0.78				0.82								0.80
Exact HWE p =			0.01				0.35								0.02
D9S303				D9S303				D9S303				D9S303			
170	4	0.50	0.19	170	2	0.25	0.16	166	1	0.25	0.25	150	2	0.14	0.10
174	4	0.50	0.19	174	3	0.38	0.18	170	1	0.25	0.25	154	1	0.07	0.07
				178	1	0.13	0.13	186	1	0.25	0.25	162	2	0.21	0.11
				182	1	0.13	0.13	190	1	0.25	0.25	166	2	0.14	0.09
				190	1	0.13	0.13					178	2	0.14	0.09
												186	2	0.14	0.09
												190	2	0.14	0.09
Alleles			2				5				4				7
Genes			8				8				4				14
Heterozygosity			0.57				0.86				1.00				0.91
Exact HWE p =			0.08				0.65				1.00				0.00

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
Western Nigeria				GGNP				NGNY				AKZN			
D11S1984				D11S1984				D11S1984				D11S1984			
164	2	0.33	0.21	124	1	0.10	0.10					104	2	0.14	0.10
188	2	0.33	0.21	128	1	0.10	0.10					144	1	0.07	0.07
192	2	0.33	0.21	164	1	0.10	0.10					164	4	0.28	0.13
				168	6	0.60	0.16					180	2	0.14	0.10
									N/a			184	2	0.14	0.10
												188	3	0.21	0.11
Alleles			3				5								6
Genes			6				10								14
Heterozygosity			0.80				0.67								0.87
Exact HWE p =			0.07				0.63								0.00
HUMPLA2A				HUMPLA2A				HUMPLA2A				HUMPLA2A			
80	1	0.17	0.17	77	1	0.08	0.08	80	1	0.25	0.25	80	2	0.13	0.08
86	2	0.33	0.21	83	5	0.42	0.15	86	3	0.75	0.25	86	12	0.75	0.11
89	3	0.50	0.22	86	4	0.33	0.14					92	2	0.13	0.08
				89	2	0.17	0.11								
Alleles			3				4				2				3
Genes			6				12				4				16
Heterozygosity			0.73				0.74				0.50				0.43
Exact HWE p =			0.19				0.27				1.00				0.01

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
Western Nigeria				GGNP				NGNY				AKZN			
D13S317				D13S317				D13S317				D13S317			
179	1	0.5	0.5	187	2	0.13	0.09					175	2	0.17	0.11
183	1	0.5	0.5	191	3	0.19	0.10					183	3	0.25	0.13
				195	3	0.19	0.10					187	2	0.17	0.11
				199	2	0.13	0.09					191	5	0.42	0.15
				215	3	0.19	0.10	N/a							
				223	1	0.06	0.06								
				227	2	0.13	0.09								
Alleles							7								4
Genes							16								12
Heterozygosity							0.90								0.77
Exact HWE p =							0.00								0.20
D16S265				D16S265				D16S265				D16S265			
78	1	0.13	0.13	80	1	0.10	0.10	106	1	0.25	0.25	80	1	0.05	0.05
80	1	0.13	0.13	82	2	0.20	0.13	108	2	0.50	0.29	90	1	0.05	0.05
96	1	0.13	0.13	106	4	0.40	0.16	114	1	0.25	0.25	100	1	0.05	0.05
102	1	0.13	0.13	108	3	0.30	0.15					104	1	0.05	0.05
108	1	0.13	0.13									106	8	0.40	0.11
106	2	0.25	0.16									108	5	0.25	0.10
110	1	0.13	0.13									110	1	0.05	0.05
												114	2	0.10	0.07
Alleles			7				4				3				8
Genes			8				10				4				20
Heterozygosity			0.96				0.78				0.83				0.79
Exact HWE p =			1.00				1.00				1.00				0.01

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
Western Nigeria				GGNP				NGNY				AKZN			
D16S539				D16S539				D16S539				D16S539			
152	7	0.70	0.15	152	12	0.67	0.11	152	5	0.63	0.18	152	10	0.45	0.11
156	2	0.20	0.13	156	2	0.11	0.08	156	1	0.13	0.13	156	2	0.09	0.06
168	1	0.10	0.10	164	4	0.22	0.10	160	1	0.13	0.13	160	4	0.18	0.08
								168	1	0.13	0.13	164	6	0.27	0.10
Alleles			3				3				4				4
Genes			10				18				8				22
Heterozygosity			0.51				0.52				0.64				0.71
Exact HWE p =			0.33				1.00				0.44				0.50
D20S470				D20S470				D20S470				D20S470			
263	1	0.17	0.17	215	2	0.20	0.13	243	1	0.25	0.25	227	1	0.07	0.07
267	2	0.33	0.21	247	1	0.10	0.10	247	1	0.25	0.25	243	1	0.07	0.07
271	1	0.17	0.17	251	1	0.10	0.10	259	1	0.25	0.25	247	1	0.07	0.07
275	1	0.17	0.17	267	2	0.20	0.13	263	1	0.25	0.25	259	2	0.13	0.09
283	1	0.17	0.17	287	1	0.10	0.10					263	3	0.19	0.10
				295	3	0.30	0.15					267	1	0.07	0.07
												271	1	0.07	0.07
												275	1	0.07	0.07
												279	1	0.07	0.07
												284	1	0.07	0.07
												287	1	0.07	0.07
												303	2	0.13	0.09
Alleles			5				6				4				12
Genes			6				10				4				16
Heterozygosity			0.93				0.89				1.00				0.96
Exact HWE p =			1.00				0.13				1.00				0.36

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
CRNP				BYM				Mosse				MTCM			
APOA2				APOA2				APOA2				APOA2			
124	1	0.05	0.05	130	2	0.20	0.13	124	1	0.08	0.08	126	1	0.20	0.13
126	1	0.05	0.05	132	8	0.8	0.13	128	1	0.08	0.08	128	1	0.10	0.10
130	5	0.25	0.10					130	3	0.25	0.13	130	1	0.10	0.10
132	11	0.55	0.11					132	2	0.17	0.11	134	1	0.10	0.10
134	2	0.10	0.07					134	2	0.17	0.11	136	2	0.20	0.13
								136	1	0.08	0.08	148	1	0.10	0.10
								138	2	0.17	0.17				
Alleles			5				2				7				6
Genes			20				10				12				7
Heterozygosity			0.65				0.36				0.91				0.93
Exact HWE p =			0.49				1.00				1.00				0.05
D4S1652				D4S1652				D4S1652				D4S1652			
122	4	0.20	0.09	122	2	0.33	0.21								
126	5	0.25	0.10	126	2	0.33	0.21								
130	2	0.10	0.07	142	2	0.33	0.21								
134	1	0.05	0.05					N/a				N/a			
138	5	0.25	0.10												
142	3	0.15	0.08												
Alleles			6				3								
Genes			20				6								
Heterozygosity			0.84				0.80								
Exact HWE p =			0.19				0.46								

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
CRNP				BYM				Mosse				MTCM			
D7S1809				D7S1809				D7S1809				D7S1809			
206	1	0.08	0.08	210	2	0.50	0.29	206	1	0.10	0.10	202		0.25	0.16
210	11	0.91	0.08	218	1	0.25	0.25	210	2	0.20	0.13	206		0.13	0.13
				22	1	0.25	0.25	214	4	0.40	0.16	210		0.38	0.18
								218	2	0.20	0.13	234		0.13	0.13
								230	1	0.10	0.10	238		0.13	0.13
Alleles			2				3				5				5
Genes			12				4				10				8
Heterozygosity			0.17				0.83				0.82				0.86
Exact HWE p =			1.00				0.34				0.07				0.84
D9S303				D9S303				D9S303				D9S303			
154	2	0.20	0.13	N/a				174	1	0.5	0.5	162	2	1.00	0.00
162	2	0.20	0.13					182	1	0.5	0.5				
170	2	0.20	0.13												
190	1	0.10	0.10												
194	3	0.30	0.15												
Alleles			5												
Genes			10												
Heterozygosity			0.87												
Exact HWE p =			0.00												
D11S1984				D11S1984				D11S1984				D11S1984			
192	2	1.00	0.00	176	2	1.00	0.00	168	2	1.00	0.00	N/a			

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
CRNP				BYM				Mosse				MTCM			
HUMPLA2A				HUMPLA2A				HUMPLA2A				HUMPLA2A			
86	4	0.67	0.21	80	2	0.20	0.13	65	2	0.13	0.09	80		0.06	0.06
95	2	0.34	0.21	86	3	0.30	0.15	80	1	0.06	0.06	86		0.44	0.12
				89	1	0.10	0.10	86	11	0.69	0.69	89		0.17	0.09
				93	1	0.10	0.10	89	1	0.06	0.06	92		0.06	0.06
				95	3	0.30	0.15	92	1	0.06	0.06	95		0.22	0.22
												98		0.06	0.06
Alleles			2				5				5				6
Genes			6				10				16				18
Heterozygosity			0.53				0.84				0.53				0.75
Exact HWE p =			0.20				0.35				0.01				0.83
D13S317				D13S317				D13S317				D13S317			
155		0.25	0.16	155		0.50	0.29	167		0.25	0.16	167		0.14	0.10
175		0.25	0.16	191		0.25	0.25	171		0.13	0.13	171		0.14	0.10
179		0.25	0.16	195		0.25	0.25	179		0.13	0.13	175		0.14	0.10
191		0.13	0.13					191		0.13	0.13	183		0.07	0.07
195		0.13	0.13					199		0.13	0.13	187		0.36	0.13
								247		0.25	0.16	191		0.14	0.10
Alleles			5				3				6				6
Genes			8				4				8				14
Heterozygosity			0.89				0.83				0.93				0.85
Exact HWE p =			0.01				0.33				0.02				0.09

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
CRNP				BYM				Mosse				MTCM			
D16S265				D16S265				D16S265				D16S265			
				84	2	0.50	0.29	82	1	0.06	0.06	80	6	0.33	0.11
				92	2	0.50	0.29	82	1	0.06	0.06	106	2	0.11	0.08
								84	1	0.06	0.06	108	6	0.33	0.11
N/a								104	1	0.06	0.06	110	3	0.17	0.09
								106	1	0.06	0.06	114	1	0.06	0.06
								108	5	0.31	0.12				
								112	2	0.13	0.09				
Alleles											7				5
Genes											16				18
Heterozygosity											0.83				0.78
Exact HWE p =											0.25				0.63
D16S539				D16S539				D16S539				D16S539			
148	1	0.05	0.05	152	5	0.63	0.18	144	1	0.06	0.06	148	1	0.05	0.05
152	13	0.65	0.11	156	1	0.13	0.13	152	9	0.50	0.12	152	11	0.50	0.11
160	1	0.05	0.05	160	1	0.13	0.13	156	3	0.17	0.09	156	4	0.18	0.08
164	5	0.25	0.25	164	1	0.13	0.13	160	2	0.11	0.08	160	1	0.05	0.05
								164	3	0.17	0.09	164	5	0.23	0.09
Alleles			4				4				5				5
Genes			20				8				18				22
Heterozygosity			0.54				0.64				0.72				0.70
Exact HWE p =			0.14				0.43				0.03				0.83

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
CRNP				BYM				Mosse				MTCM			
D20S470				D20S470				D20S470				D20S470			
203	7	0.50	0.14	203	2	0.25	0.16	215	1	0.08	0.08	203	1	0.07	0.07
227	1	0.07	0.07	247	1	0.13	0.13	227	2	0.17	0.11	207	1	0.07	0.07
255	1	0.07	0.07	259	1	0.13	0.13	259	2	0.17	0.11	215	1	0.07	0.07
263	3	0.21	0.11	263	3	0.38	0.10	263	2	0.17	0.11	259	2	0.14	0.10
279	2	0.14	0.10	283	1	0.13	0.13	279	2	0.17	0.11	263	2	0.14	0.10
								283	1	0.08	0.11	279	1	0.07	0.07
								287	2	0.17	0.08	283	5	0.36	0.13
											0.11	287	1	0.07	0.07
Alleles			5				5				7				8
Genes			14				8				12				14
Heterozygosity			0.73				0.86				0.93				0.87
Exact HWE p =			0.01				0.08				0.14				0.87
Man				DEFR				Campo				DBR			
APOA2				APOA2				APOA2				APOA2			
120	2	0.11	0.08	128	1	0.06	0.06	126	4	0.18	0.08	130	1	0.06	0.06
126	3	0.17	0.09	130	2	0.11	0.08	130	3	0.14	0.07	132	7	0.39	0.12
128	2	0.11	0.08	132	8	0.44	0.12	132	8	0.36	0.10	134	4	0.22	0.10
130	4	0.22	0.10	134	4	0.22	0.10	134	5	0.22	0.09	136	2	0.11	0.08
132	6	0.33	0.11	136	1	0.06	0.06	136	1	0.05	0.05	140	4	0.22	0.10
134	1	0.06	0.06	140	2	0.11	0.08	140	1	0.05	0.05				
Alleles			6				6				6				5
Genes			18				18				22				18
Heterozygosity			0.83				0.77				0.79				0.78
Exact HWE p =			0.12				0.97				0.00				0.27

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
Man				DEFR				Campo				DBR			
D4S1652				D4S1652				D4S1652				D4S1652			
118	3	0.30	0.15	118	1	0.13	0.13	118	3	0.17	0.09	122	1	0.17	0.17
122	6	0.60	0.16	122	3	0.36	0.18	122	3	0.17	0.09	142	2	0.33	0.22
150	1	0.10	0.10	138	2	0.25	0.16	126	1	0.06	0.06	146	3	0.50	0.22
				150	1	0.13	0.13	130	4	0.22	0.10				
				158	1	0.13	0.13	134	2	0.11	0.08				
								138	1	0.06	0.06				
								142	1	0.06	0.06				
								146	2	0.17	0.09				
								150	1	0.06	0.06				
Alleles			3				5				10				3
Genes			10				8				18				6
Heterozygosity			0.60				0.86				0.92				0.73
Exact HWE p =			1.00				0.09				0.00				0.20
D7S1809				D7S1809				D7S1809				D7S1809			
206	1	0.13	0.13	210	6	0.50	0.15	218	1	0.13	0.13	210	2	0.33	0.17
210	1	0.13	0.13	226	4	0.33	0.14	230	2	0.25	0.16	222	1	0.17	0.17
214	2	0.25	0.25	238	1	0.08	0.08	238	3	0.38	0.18	242	3	0.50	0.22
218	3	0.38	0.38	242	1	0.08	0.08	242	2	0.25	0.16				
242	1	0.13	0.13												
Alleles			5				4				4				3
Genes			8				12				8				6
Heterozygosity			0.86				0.68				0.77				0.73
Exact HWE p =			0.09				0.01				0.78				1

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
Man				DEFR				Campo				DBR			
D9S303				D9S303				D9S303				D9S303			
154	1	0.10	0.10	142	1	0.07	0.07	146	1	0.08	0.08	150	2	0.14	0.10
158	2	0.20	0.13	146	4	0.28	0.13	166	3	0.25	0.13	158	1	0.07	0.07
162	3	0.30	0.15	154	1	0.07	0.07	170	6	0.50	0.15	162	1	0.07	0.07
166	1	0.10	0.10	162	2	0.14	0.10	174	1	0.08	0.08	170	6	0.42	0.13
170	3	0.30	0.10	170	1	0.07	0.07	186	1	0.08	0.08	174	1	0.07	0.07
				174	1	0.07	0.07					178	1	0.07	0.07
				178	2	0.14	0.10					182	1	0.07	0.07
				182	2	0.14	0.10					186	1	0.07	0.07
Alleles			5				8				5				8
Genes			10				14				12				14
Heterozygosity			0.84				0.90				0.72				0.82
Exact HWE p =			0.36				0.21				1.00				0.00
D11S1984				D11S1984				D11S1984				D11S1984			
180	2	1.00	0.00	168	2	0.20	0.13	176	2	0.25	0.16	180	2	0.20	0.13
				176	1	0.10	0.10	180	2	0.25	0.16	184	4	0.40	0.16
				180	2	0.20	0.13	184	3	0.38	0.18	188	4	0.40	0.16
				184	3	0.30	0.15	196	1	0.13	0.13				
				192	1	0.10	0.10								
				196	1	0.10	0.10								
Alleles							6				4				3
Genes							10				8				10
Heterozygosity							0.89				0.82				0.71
Exact HWE p =							0.01				0.03				0.01

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
Man				DEFR				Campo				DBR			
HUMPLA2A				HUMPLA2A				HUMPLA2A				HUMPLA2A			
77	4	0.29	0.13	68	1	0.05	0.05	83	3	0.30	0.16	86	4	0.50	0.19
86	7	0.50	0.14	86	14	0.70	0.11	86	5	0.50	0.17	89	2	0.25	0.16
95	3	0.21	0.11	89	3	0.15	0.08	98	2	0.20	0.13	98	2	0.25	0.16
				98	2	0.10	0.07								
Alleles			3				4				3				3
Genes			14				20				10				8
Heterozygosity			0.67				0.50				0.67				0.71
Exact HWE p =			0.01				0.03				0.05				0.03
D13S317				D13S317				D13S317				D13S317			
167	2	0.14	0.10	175	2	0.20	0.13	179	4	0.33	0.21	179	2	0.33	0.21
175	3	0.21	0.11	187	1	0.10	0.10	183	2	0.17	0.17	183	1	0.17	0.17
179	2	0.14	0.10	191	3	0.30	0.15	187	4	0.33	0.21	187	2	0.33	0.21
183	1	0.07	0.07	199	1	0.10	0.10	191	2	0.17	0.17	191	1	0.17	0.17
187	1	0.07	0.07	203	1	0.10	0.10								
191	2	0.14	0.10	251	1	0.10	0.10								
195	1	0.07	0.07	255	1	0.10	0.10								
247	1	0.07	0.07												
255	1	0.07	0.07												
Alleles			9				7				4				4
Genes			14				10				12				6
Heterozygosity			0.93				0.91				0.77				0.87
Exact HWE p =			0.00				0.04				0.09				0.07

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
Man				DEFR				Campo				DBR			
D16S265				D16S265				D16S265				D16S265			
80	3	0.17	0.09	80	5	0.36	0.13	94	1	0.17	0.17	78	1	0.07	0.07
82	2	0.11	0.08	106	3	0.21	0.11	102	1	0.17	0.17	80	1	0.07	0.14
84	1	0.06	0.06	108	6	0.42	0.14	108	1	0.17	0.17	106	6	0.43	0.14
90	1	0.06	0.06					110	1	0.17	0.17	108	5	0.36	0.13
106	5	0.28	0.11					114	2	0.33	0.21	110	1	0.07	0.07
108	6	0.33	0.11												
Alleles			6				3				5				5
Genes			18				14				6				14
Heterozygosity			0.81				0.69				0.93				0.73
Exact HWE p =			0.30				0.48				0.21				0.33
D16S539				D16S539				D16S539				D16S539			
148	3	0.11	0.06	152	12	0.67	0.11	152	11	0.55	0.11	148	1	0.05	0.05
152	12	0.43	0.10	156	1	0.05	0.06	156	1	0.05	0.05	152	16	0.72	0.10
156	4	0.14	0.07	164	5	0.28	0.11	160	1	0.05	0.05	156	2	0.10	0.06
164	9	0.32	0.09					164	7	0.35	0.11	164	3	0.14	0.07
Alleles			5				3				4				4
Genes			28				18				20				22
Heterozygosity			0.72				0.50				0.60				0.46
Exact HWE p =			0.03				0.63				0.05				0.40

Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.	Allele	Count	%	S.D.
Man				DEFR				Campo				DBR			
D20S470				D20S470				D20S470				D20S470			
203	3	0.14	0.06	203	1	0.06	0.06	167	2	0.14	0.09	203	3	0.15	0.08
211	1	0.05	0.05	207	1	0.06	0.06	215	2	0.14	0.10	215	1	0.05	0.05
215	1	0.05	0.05	255	1	0.06	0.06	247	2	0.14	0.10	247	2	0.10	0.07
247	1	0.05	0.05	259	2	0.13	0.13	251	1	0.07	0.07	255	2	0.10	0.07
263	2	0.09	0.06	263	4	0.25	0.11	263	1	0.07	0.07	259	4	0.20	0.10
267	1	0.05	0.05	267	1	0.06	0.06	267	3	0.21	0.11	263	1	0.05	0.05
275	2	0.09	0.09	283	4	0.25	0.11	279	2	0.14	0.10	275	2	0.10	0.07
279	1	0.05	0.05	287	2	0.13	0.09	283	1	0.07	0.07	279	2	0.10	0.07
283	7	0.31	0.10									283	3	0.15	0.09
287	2	0.09	0.06												
291	1	0.05	0.05												
Alleles			11				8				8				9
Genes			22				16				14				20
Heterozygosity			0.88				0.88				0.92				0.92
Exact HWE p =			0.64				0.10				0.00				0.01

BIBLIOGRAPHY

Agbelusi, E. A. 1994. Wildlife conservation in Ondo State. *The Nigerian Field* 59: 73-83.

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

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

Amos. W. and D. C. Rubinstzein. 1996. Microsatellites are subject to directional evolution. *Nature Genetics* 12: 13-14.

Amos. W., S. J. Sawcer, R. W. Feakes. and D. C. Rubinsztein. 1996. Microsatellites show mutational bias and heterozygote instability. *Nature Genetics* 13: 390-391.

Anderson. S., A. T. Bankier. B. G. Barell. M. H. L. d. Bruijn, A. R. Coulson. J. Drouin. I. C. Eperson, D. P. Nierlich. B. A. Roe, 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.

Aris-Brosou, S. and L. Excoffier. 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Molecular Biology and Evolution* 13: 494-504.

Aubréville, A. 1937. Les Forets du Dahomey et du Togo. *Bulletin Comité d'Etudes Histoire et Scientifique Afrique Occidentale* 20: 1-112.

Aubréville, A. 1949. Climats, Forets et Desertification de l'Afrique Tropicale:175-187. Paris: Societe d'Editions Geographiques, Maritimes et Coloniales.

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

Avise, J. C. 1989b. A role for molecular genetics in the recognition and conservation of endangered species. *Trends in Ecology and Evolution* 4: 279-284.

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

Avise, J. C. 1998. The history and purview of phylogeography: a personal reflection. *Molecular Ecology* 7: 371-379.

Avise, J. C. and J. L. Hamrick. 1996. *Conservation Genetics : Case Histories from Nature.* New York: Chapman & Hall.

Ayala, F. J. and A. A. Escalante. 1996. The evolution of human populations: a molecular perspective. *Molecular Phylogenetics Evolution* 5: 188-201.

Baker, R. J. 1994. Some thoughts on conservation, biodiversity, museums, molecular characters, systematics and basic research. *Journal of Mammalogy* 75: 277-287.

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

Barnwell. R. 1991. Gashaka-Gumti Game Reserve Gongola State Nigeria: An Indicative Proposal for Developing Gashaka-Gumti as a National Park. U.K.: WWF.

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

Bertorelle, G. and M. Slatkin. 1995. The number of segregating sites in expanding human populations, with implications for estimates of demographic parameters. *Molecular Biology Evolution* 12: 887-892.

Blanquer-Maumont, A. and B. Crouau-Roy. 1995. Polymorphism, monomorphism, and sequences in conserved microsatellites in primate species. *Journal of Molecular Evolution* 41: 492-497.

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.

Booth, A. H. 1958a. The Niger, the Volta and the Dahomey Gap as geographic barriers. *Evolution* 12: 48-62.

Booth, A. H. 1958b. The zoogeography of West African primates: A review. *Bulletin de l'I.F.A.N.* A: 587-622.

Bossart, J. L. and D. P. Powell. 1998. Genetic estimates of population structure and gene flow: Limitations, lessons and new directions. *Trends in Ecology and Evolution* 13: 202-206.

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: 455-457.

Braemer, K. 1992. Ancestral Areas: A cladistic reinterpretation of the center of origin concept. *Systematic Biology* 41: 436-445.

Bremer, K. 1994. Branch support and tree stability. *Cladistics* 10: 295-304.

Brinkmann, B., M. Klintschar, F. Neuhuber, J. Hühne, and B. Rolf. 1998. Mutation rate in microsatellites: Influence of the structure and length of the tandem repeat. *American Journal of Human Genetics* 62: 1408-1415.

Brown, J. H. and A. C. Gibson. 1983. *Biogeography*. Oxford: C.V. Mosby.

Brown, W. M. 1983. Evolution of animal mitochondrial DNA. In *Evolution of Genes and Proteins*, ed. M. Nei and R. K. Koehn:62-88. Sunderland, MA: Sinauer Associates, Inc.

Bruford, M. W. and R. K. Wayne. 1993. Microsatellites and their application to population genetic studies. *Current Opinions in Genetics and Development* 3: 939-943.

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

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

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

Chakraborty, R., M. Kimmel, D. N. Stivers, L. J. Davison, and R. Deka. 1997. Relative mutation rates at di-, tri-, and tetranucleotide microstallite loci. *Proceedings of the National Academy of Sciences USA* 94: 1041-1046.

Chapman, C. A. 1983. Speciation of tropical rainforest primates of Africa: Insular biogeography. *African Journal of Ecology* 21: 297-308.

Chapman, C. A. and R. W. Wrangham. 1993. Range use of the forest chimpanzees of Kibale: Implications for the understanding of chimpanzee social organization. *American Journal of Primatology* 31: 262-273.

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

Cockerham, C. C. and B. S. Weir. 1993. Estimation of gene flow from F-statistics. *Evolution* 47: 855-863.

Collins, D. A. and W. C. McGrew. 1988. Habitats of three groups of chimpanzees (*Pan troglodytes*) in western Tanzania compared. *Journal of Human Evolution* 17: 552-574.

Collura, R. V. and C. B. Stewart. 1995. Insertions and duplications of mtDNA in the nuclear genome of old world monkeys and hominoids. *Nature* 378: 485-489.

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

Conway, H., B. L. Hall, G. H. Denton, A. M. Gades, and E. D. Waddington. 1999. Past and future grounding-line retreat of the West Antarctic Ice Sheet. *Science* 286: 280-283.

Coolidge, H. J. 1933. *Pan paniscus*. Pigmy chimpanzee from south of the Congo River. *American Journal of Physical Anthropology* 18: 1-57.

Cooper, G., D. C. Rubinsztein, and W. Amos. 1998. Ascertainment bias cannot entirely account for human microsatellites being longer than their chimpanzee homologues. *Human Molecular Genetics* 7: 1425-1429.

Cox, C. B. and P. D. Moore. 1993. *Biogeography: An Ecological and Evolutionary Approach*. New York: Blackwell Scientific Publications.

Crandall, K. A. and A. R. Templeton. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* 134: 959-969.

Crawford, A. M., S. M. Kappes, K. A. Paterson, M. J. deGotari, K. G. Dodds, B. A. Freking, R. T. Stone, and C. W. Beattie. 1998. Microsatellite evolution: testing the ascertainment bias hypothesis. *Journal of Molecular Evolution* 46: 256-260.

Crouau-Roy, B., S. Service, M. Slatkin, and N. and Freimer. 1996. A fine-scale comparison of the human and chimpanzee genomes--linkage, linkage disequilibrium and sequence-analysis. *Human Molecular Genetics* 5: 1131-1137.

Crozier, R. H. 1992. Genetic diversity and the agony of choices. *Biological Conservation* 61: 11-15.

Deinard, A., R. Dorit, C. Castiglione, Z. Jiang, D. Becker, F. Ruddle, K. Schugart, and K. Kidd. 1999. Evolution of the HOXB6 intergenic region: motif conservation at the lateral plate mesoderm (LPM) enhancer element. *Journal of Experimental Zoology* 285: 170-176.

Deinard, A. and K. Kidd. 1996. Identifying "wild" levels of genetic diversity within *Pan paniscus* and *Pan troglodytes*. In *16th Congress of the International Primatological Society*:784. Madison, WI.

Deinard, A. and K. Kidd. 1999. Evolution of a HOXB6 intergenic region within the great apes and humans. *Journal of Human Evolution* 36: 687-703.

Deinard, A. S. and K. Kidd. 2000. Identifying conservation units within captive chimpanzee populations. *American Journal of Physical Anthropology* 111: 25-44.

deMenocal, P. B. 1995. Plio-Pleistocene African climate. *Science* 270: 53-59.

DeSalle, R., C. Wray, and R. Abasher. 1994. Computational problems in molecular systematics. In *Molecular Ecology and Evolution: Approaches and Applications*, ed. B. Scwater, B. Stredt, G. P. Wagner, and R. DeSalle:353-370. Basel: Verlag.

Di Rienzo, A., A. C. Peterson, J. C. Garza, A. M. Valdes, M. Slatkin, and N. B. Freimer. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences U S A* 91: 3166-3170.

Domingo-Roura, X., T. Lopez-Giraldez, M. Shinohara, and O. Takenaka. 1997. Hypervariable microsatellite loci in the Japanese macaque (*Macaca fuscata*) conserved in related species. *American Journal of Primatology* 43: 357-360.

Don, R. H., P. T. Cox, B. J. Wainwright, K. Baker, and J. S. Mattick. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* 19: 4008.

Dowsett, R. J., ed. 1989. *A Preliminary Natural History Survey of Mambilla Plateau and Some Lowland Forests of Eastern Nigeria*: Tauraco Research Report 1.

Dowsett, R. J. and F. Dowsett-Lemaire. 1989. Larger mammals in the Gotel Mts and on the Mambilla Plateau, eastern Nigeria. In *A Preliminary Natural History Survey of Mambilla Plateau and Some Lowland Forests of Eastern Nigeria*, ed. R. J. Dowsett:38-41.

Doyle, J. J., M. Morgante, S. V. Tingey, and W. Powell. 1998. Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (Glycine subgenus Glycine). *Molecular Biology and Evolution* 15: 215-218.

Dunn, A. 1993. Large Mammal Survey of Gashaka-Gumti National Park. U.K: WWF.

Edwards, A., H. A. Hammond, L. Jin, C. T. Caskey, and R. Chakroborty. 1992. Genetic variaiton at five trimeric and tetrameric tandem repeat loci in four human groups. *Genomics* 12: 241-253.

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

Elliot, D. G. 1913. Review of the Primates. *American Museum of Natural History Monograph* 1.

Ely, J. J., D. L. Gonzalez, A. Reeves-Daniel, and W. H. Stone. 1998. Individual identification and paternity determination in chimpanzees (*Pan troglodytes*) using human short tandem repeat (STR) markers. *International Journal of Primatology* 19: 255-264.

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

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

Ern, H. 1988. Flora and Vegetation of the Dahomey Gap--A Contribution to the Plant Geography of West Tropical Africa. *Monographs of the Missouri Botanical Society* 25: 517-520.

Estoup, A., L. Garnery, M. Solignac, and J. M. Cornuet. 1995. Microsatellite variation in honey bee (*Apis mellifera*) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics* 140: 679-695.

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

Fahrig, L. and G. Merriam. 1994. Conservation of fragmented populations. *Conservation Biology* 8: 50-59.

Faith, D. P. 1994. Genetic diversity and taxonomic priorities for conservation. *Biological Conservation* 68: 69-74.

Felsenstein, J. 1993. PHYLIP: Phylogenetic Inference Package 3.5. Department of Genetics, University of Washington, Seattle.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17: 368-376.

Felsenstein, J. 1984. Distance methods for inferring phylogenies: A justification. *Evolution* 38: 16-24.

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

Fischer, H., M. Wahlen, J. Smith, D. Mastroianni, and B. Deck. 1999. Ice core records of atmospheric CO2 around the last three glacial terminations. *Science* 283: 1712-1714.

Fluckiger, J., A. Dallenbach, T. Blunier, B. Stauffer, T. F. Stocker, D. Raynaud, and J. Barnola. 1999. Variations in atmospheric N_2O concentration during abrupt climatic changes. *Science* 285: 227-230.

Fruth, B. and G. Hohmann. 1996. Nest building behavior in the great apes: the great leap forward. In *Great Ape Societies*, ed. W. C. McGrew, L. F. Marchant, and T. Nishida:225-240. Cambridge: Cambridge University Press.

Fukada, M., S. Wakasugi, and T. Tsuzuki. 1985. Mitochondrial-like DNA sequences in the human nuclear genome. *Journal of Molecular Biology* 186: 257-266.

Gadsby, E. L. and P. D. Jenkins. 1992. Report on Wildlife and Hunting in the Proposed Etinde Forest Reserve. Limbe: Limbe Botanic Garden & Rainforest Genetic Conservation Project.

Gagneux, P. 1998. Population Genetics of West African Chimpanzees (*Pan troglodytes verus*), University of Basel.

Gagneux, P., C. Boesch, and D. S. Woodruff. 1997. Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* 6: 861-868.

Gagneux, P., C. Boesch, and D. S. Woodruff. 1999. Female reproductive strategies and community structure in wild West African chimpanzees. *Animal Behaviour* 57: 19-32.

Gagneux, P., C. Wills, U. Gerloff, D. Tautz, P. A. Morin, C. Boesch, B. Fruth, G. Hohmann, O. A. Ryder, and D. S. Woodruff. 1999. Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proceedings of the National Academy of Sciences U S A* 96: 5077-5082.

Gagneux, P., D. S. Woodruff, and C. Boesch. 1997. Furtive mating in female chimpanzees. *Nature* 387: 358-359.

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

Garner, K. J. and O. A. Ryder. 1996. Mitochondrial DNA diversity in gorillas. *Molecular Phylogenetics and Evolution* 6: 39-48.

Gartlan, J. S. and T. T. Struhsaker. 1972. Polyspecific associations and niche separation of rain-forest anthropoids in Cameroon, West Africa. *Journal of the Zoological Society of London* 168: 221-266.

Garza, J. C. and N. B. Freimer. 1996. Homoplasy for size at microsatellite loci in humans and chimpanzees. *Genome Research* 6: 211-217.

Garza, J. C., M. Slatkin, and N. B. Freimer. 1995. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Molecular Biology and Evolution* 12: 594-603.

Giles, B. E. 1994. Genetic biodiversity: Analysing the data. *Trends in Ecology and Evolution* 9: 317-319.

Gilpin, M. E. and M. E. Soulé. 1986. Minimum viable populations: The processes of species extinction. In *Conservation Biology: The Science of Scarcity and Diversity*, ed. M. E. Soule:13-34. Sunderland, MA: Sinauer Associates, Inc.

Givnish, T. J. and K. J. Sytsma. 1997. Consistency, characters, and the likelihood of correct phylogenetic inference. *Molecular Phylogenetics and Evolution* 7: 320-330.

Goldberg, T. L. 1996. Genetics and Biogeography of East African Chimpanzees *(Pan troglodytes schweinfurthii*). Ph.D., Harvard University.

Goldberg, T. L. 1998. Biogeographic predictors of genetic diversity in populations of East African chimpanzees (*Pan troglodytes schweinfurthii*). *International Journal of Primatology* 19: 237-254.

Goldberg, T. L. and M. Ruvolo. 1997. The geographic apportionment of mitochondrial genetic diversity in east African chimpanzees, *Pan troglodytes schweinfurthii*. *Molecular Biology and Evolution* 14: 976-984.

Goldman, N. 1990. Maximum likelihood of phylogenetic trees, with special reference to Poisson process models of DNA substitution and to parsimony analysis. *Systematic Zoology* 39: 345-361.

Goldstein, D. B., A. R. Linares, M. W. Feldman, and L. L. Cavalli-Sforza. 1995a. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139: 463-471.

Goldstein, D. B. and D. D. Pollock. 1994. Least squares estimation of molecular distance--noise abatement in phylogenetic reconstruction. *Theoretical Population Biology* 45: 219-226.

Goldstein, D. B. and D. D. Pollock. 1997. Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* 51: 27-47.

Goldstein, D. B., G. W. Roemer, D. A. Smith, D. E. Reich, A. Bergmen, and R. K. Wayne. 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. *Genetics* 151: 797-801.

Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995b. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences U S A* 92: 6723-6727.

Goloboff, P. 1999. NONA 2.0. Fundacion e Instituto Miguel Lillo, Tucuman, Argentina.

Gonder, M. K., J. F. Oates, T. R. Disotell, M. R. Forstner, J. C. Morales, and D. J. Melnick. 1997. A new west African chimpanzee subspecies? *Nature* 388: 337.

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

Goodman, M. and R. E. Tashian. 1969. A geographic analysis in the serum transferrin and red cell phosphoglucomutase polymorphisms of chimpanzees. *Human Biology* 41: 237-249.

Goodman, M., W. G. Wisecup, H. H. Reynolds, and C. H. Kratochvil. 1967. Transferrin polymorphism and population differences in the genetic variability of chimpanzees. *Science* 156: 98-100.

Goodman, S. J. 1997. Rst CALC: A collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology* 6: 881-885.

Goosens, B., L. P. Waits, and P. Taberlet. 1998. Plucked hair samples as a source of DNA: Reliability of dinucleotide microsatellite genotyping. *Molecular Ecology* 7: 1237-1241.

Gray, J. E. 1862. List of mammalia from the Cameroon mountains. collected by Captain Burton. *Proceedings of the Zoological Society of London*: 180-181.

Gray, M. W., G. Burger, and B. F. Lang. 1999. Mitochondrial evolution. *Science* 283: 1476-1481.

Grimaldi, M. and B. Crouau-Roy. 1997. Microsatellite allelic homoplasy due to variable flanking regions. *Journal of Molecular Evolution* 44: 336-340.

Groves, C. P. 1993. Order Primates. In *Mammalian Species of the World: A Taxonomic and Geographic Reference*, ed. D. E. Wilson and D. M. Reader: 243-277. Washington, D.C.: Smithsonian Institution Press.

Groves, C. P., C. Westwood, and B. T. Shea. 1993. Unfinished business: Mahalanobis and a clockwork orang. *Journal of Human Evolution* 22: 22-37.

Grubb, P. 1982. Refuges and dispersal in the speciation of African forest mammals. In *Biological Diversification in the Tropics*, ed. G. T. Prance:537-553. New York: Academic Press.

Grubb, P. 1990. Primate geography in the Afro-tropical forest biome. In *Vertebrates in the Tropics*, ed. G. Peters and R. Hutterer:187-214. Bonn: Museum Alexander Koenig.

Guo, S. W. and E. A. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361-372.

Hacker, J. E., G. Cowlishaw, and P. H. Williams. 1998. Patterns of African primate diversity and their evaluation for the selection of conservation areas. *Biological Conservation* 84: 251-262.

Haffer, J. 1969. Speciation in Amazon forest birds. *Science* 165: 131-137.

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

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

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

Hamilton, A. C. 1992. History of forests and climate. In *Conservation Atlas of Tropical Forests, Africa*, ed. J. Sayer, C. S. Harcourt, and N. M. Collins:17-25. New York: MacMillian.

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

Harpending, H. C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* 66: 591-600.

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

Harris, E. E. and J. Hey. (1999). X chromosome evidence for ancient human histories. *Proceedings of the National Academy of Sciences U S A* 96: 3320-3324.

Hasegawa, M. and M. Fujiwara. 1993. Relative efficiencies of the maximum likelihood, maximum parsimony, and neighbor-joining methods for estimating protein phylogeny. *Molecular Phylogenetics and Evolution* 2: 1-5.

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

Hey, J. 1997. Mitochondrial and nuclear genes present conflicting portraits of human origins. *Molecular Biology and Evolution* 14: 166-172.

Highuchi, R., D. H. v. Beroldigen, G. F. Sensabaugh, and H. A. Erlich. 1988. DNA typing from single hairs. *Nature* 332: 543-546.

Hill, W. C. O. 1967. The taxonomy of the genus Pan. In *Progress in Primatology*, ed. D. Stark, D. Schneider, and H. Kuhn:47-54. Stutgart: Fischer.

Hill, W. C. O. 1969. The nomenclature, taxonomy and distribution of chimpanzees. In *The Chimpanzee*, ed. G. H. Bourne, 1:22-49. Basel: Karger.

Hillis, D. M. 1995. Approaches for assessing phylogenetic accuracy. *Systematic Biology* 44: 3-16.

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.

Hillis, D. M., B. K. Mable, and C. Moritz. 1996. Applications of molecular systematics: The state of the field and a look to the future. In *Molecular Systematics*, ed. D. M. Hillis, B. K. Mable, and C. Moritz:515-543. Sunderland, MA: Sinauer Associates, Inc.

Hillis, D. M., C. Moritz, and B. K. Mable, eds. 1996. *Molecular Systematics*. Sunderland, MA: Sinaeur Associates, Inc.

Ho, K. F., E. M. Craddock, F. Piano, and M. P. Kambysellis. 1996. Phylogenetic analysis of DNA length mutations in a repetitive region of Hawaiin *Drosophila* yolk protein gene *Yp2*. *Journal of Molecular Evolution* 43: 116-124.

Hoelzer, G. A., J. Wallman, and D. J. Melnick. 1998. The effects of social structure, geographical structure, and population size on the evolution of mitochondrial DNA: II. Molecular clocks and the lineage sorting period. *Journal of Molecular Evolution* 47: 21-31.

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

Howell, N., I. Kubacka, and D. A. Mackey. 1996. How rapidly does the human mitochondrial genome evolve? *American Journal of Human Genetics* 59: 501-509.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. In *Oxford Surveys in Evolutionary Biology*, ed. D. J. Futuyma and J. Antonovics:1-44. Oxford: Oxford University Press.

Hudson, R. R., M. Slatkin, and W. P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132: 583-589.

Huelsenbeck, J. P. 1995. The robustness of two phylogenetic methods: Four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor-joining. *Molecular Biology and Evolution* 12: 843-849.

Imbrie, J. 1985. A theoretical framework for the Pleistocene ice ages. *Journal of the Geological Society of London* 142: 417-432.

Imbrie, J., E. A. Boyle, S. C. Clemens, A. Duffy, W. R. Howard, J. Kukla, D. G. Kutzbach, G. Martinson, A. MacIntyre, C. Mix, B. Molfino, J. J. Morely, L. C. Peterson, W. L. Pisias, W. L. Prell, M. E. Raymo, N. J. Shackleton, and J. R. Toggweiler. 1992. On the structure and origin of major glacial cycles 1. Linear responses to Milankovitch forcing. *Paleooceanography* 7: 701-738.

Imbrie, J. and J. Z. Imbrie. 1980. Modeling the climatic response to orbital variations. *Science* 207: 943-953.

Ina, Y. 1998. Estimation of the transition/transversion ratio. *Journal of Molecular Evolution* 46: 521-533.

Innis, M. A., D. H. Gelfland, J. J. Sninsky, and T. J. White, eds. 1990. *PCR Protocols: A Guide to Methods and Applications*. San Diego: Academic Press.

Jarne, P. and J. L. Lagoda. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* 11: 424-429.

Jin, L., C. Macuabas, J. Hallmayer, A. Kimura, and E. Mignot. 1996. Mutation rate varies among alleles at a microsatellite locus: Phylogenetic evidence. *Proceedings of the National Academy of Sciences, USA* 93: 15285-15288.

Jin, L., P. A. Underhill, M. Buoncristiani, and J. M. Robertson. 1997. Defining microsatellite alleles by genotyping global indigenous human populations and non-human primates. *Journal of Forensic Sciences* 42: 496-499.

Johnston, H. H. 1905. On the nomenclature of the anthropoid apes as proposed by the Hon. Walter Rothschild. *Proceedings of the Zoological Society of London* 2: 70-74.

Jolly, C. J., J. F. Oates, and T. R. Disotell. 1995. Chimpanzee kinship. *Science* 268: 185-186.

Jones, A. G., G. Rosenqvist, A. Berglund, and J. C. Avise. 1999. Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. *Genetics* 152: 1057-1063.

Jorde, L. B., M. Bamshad, and A. R. Rogers. 1998. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* 20: 126-136.

Kaessmann, H., V. Wiebe, and S. Pääbo. 1999. Extensive nuclear DNA sequence diversity among chimpanzees. *Science* 286: 1159-1162.

Kimura, M. 1969. The number of heterozygous nucleotide sites maintained in a finite populations due to a steady flux of mutations. *Genetics* 61: 893-603.

Kingdon, J. 1989. *Island Africa*. Princeton: Princeton University Press.

Kishino, H. and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution* 29: 170-179.

Kocher, T. D. and A. C. Wilson. 1991. Sequence evolution of mitochondrial DNA in human and chimpanzees. In *Evolution of Life: Fossils, Nucleotides and Culture*:391-413. Tokyo: Springer.

Koenig, W. D. 1999. Spatial autocorrelation of ecological phenomenon. *Trends in Ecology and Evolution* 14: 22-26.

Kortlandt, A. 1983. Marginal habitats of chimpanzees. *Journal of Human Evolution* 12: 231-278.

Kortlandt. A. 1991. Open letter.

Kuhner, M. K. and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution* 11: 459-468.

Lang, C., M. Leuenberger, J. Schwander, and S. Johnsen. 1999. 16 degrees C rapid temperature variation in central Greenland 70,000 years Ago. *Science* 286: 934-937.

Laurenson, M. K. and T. M. Caro. 1994. Ecological and genetic factors in conservation: A cautionary tale. *Science* 263: 485-486.

Lee, P. C., J. Thornback, and E. L. Bennett. 1988. The Threatened Primates of Africa: The IUCN Red Data Book. Gland, Switzerland: IUCN.

Li, W. H. and D. Graur. 1991. *Fundamentals of Molecular Evolution*. Sunderland, MA: Sinauer Associates, Inc.

Livingstone, D. A. 1975. Late Quaternary climatic change in Africa. *Annual Review of Ecology and Systematics* 6: 249-280.

Livingstone, D. A. 1993. Evolution of African climate. In *Biological Relationships between Africa and South America*, ed. P. Goldblatt:455-472. New Haven: Yale University Press.

Lowell, T. V., C. J. Heusser, B. G. Andersen, P. I. Moreno, A. Hauser, L. E. Heusser, C. Schluchter, D. R. Marchant, and G. H. Dento. 1995. Interhemispheric correlation of late Pleistocene glacial events. *Science* 269: 154-162.

Lynch, M. and T. J. Crease. 1990. The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution* 7: 377-394.

Maddison, D. R., M. Ruvolo, and D. L. Swofford. 1992. Geographic origins of human mitochondrial DNA: phylogenetic evidence from control region sequences. *Systematic Biology* 41: 111-124.

Maddison, W. P. and D. R. Maddison. 1996. MacClade: Analysis of Phylogeny and Character Evolution 3.6. Sinauer Associates, Inc., Sunderland, MA.

Maley, J. 1991. The African rainforest vegetation and paleoenvironments during the late Quaternary. *Climatic Change* 19: 79-98.

Maley, J. 1996. The African rain forest--main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. *Proceedings of the Royal Society of Edinburgh* 104B: 31-73.

Maley, J., D. A. Livingstone, P. Giresse, N. Thouveny, P. Brenac, K. Kelts, G. Kling, C. Stager, M. Haag, M. Fournier, Y. Bandet, D. Williamson, and A. Zogning. 1990. Lithostratigraphy, volcanism, paleomagnetism and palynology of Quaternary lacrustrine deposits from Barombi Mbo (West Cameroon): preliminary results. *Journal of Volcanology and Geothermal Research* 42: 319-335.

Marchesi, P., N. Marchesi, B. Fruth, and C. Boesch. 1995. Census and distribution of chimpanzees in Cote d'Ivoire. *Primates* 36: 591-607.

Marjoram, P. and P. Donnelly. 1994. Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications of early human evolution. *Genetics* 136: 673-683.

Mayewski, P. A., M. S. Twickler, S. I. Whitlow, L. D. Meeker, Q. Yang, J. Thomas, K. Kreutz, P. M. Grootes, D. L. Morse, E. J. Steig, E. D. Waddington, E. S. Saltzman, P. Whung, and K. C. Taylor. 1996. Climate change during the last deglaciation in Antarctica. *Science* 272: 1636-1638.

Mayr, E. and R. J. O'Hara. 1986. The biogeographic evidence supporting the Pleistocene forest refuge hypothesis. *Evolution* 40: 55-67.

McAllister, B. F. and J. H. Werren. 1999. Evolution of tandemly repeated sequences: What happens at the end of an array? *Journal of Molecular Evolution* 48: 469-481.

McGrew, W. C., L. F. Marchant, and T. Nishida. 1996. *Great Ape Societies.* Cambridge ; New York, NY, USA: Cambridge University Press.

Meffe, G. K. and C. R. Carroll. 1994. *Principles of Conservation Biology*. Sunderland, MA: Sinauer Asscoiates, Inc.

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. *Journal of Molecular Evolution* 33: 543-555.

Metzgar, D., D. Field, R. Haubrich, and C. Wills. 1998. Sequence analysis of a compound coding-region microsatellite in *Candida albicans* resolves homoplasies and provides a high-resolution tool for genotyping. *FEMS Immunology Medicine and Microbiology* 20: 103-109.

Meyer, E., P. Wiegand, S. P. Rand, D. Kuhlmann, M. Brack, and B. Brinkmann. 1995. Microsatellite polymorphisms reveal phylogenetic relationships in primates. *Journal of Molecular Evolution* 41: 10-14.

Meyer, S., G. Weiss, and A. von Haesler. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152: 1103-1110.

Michalakis, Y. and L. Excoffier. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellites. *Genetics* 142: 1061-1064.

Minch, E., A. Ruiz-Lanares, D. Goldstein, M. Feldman, and L. L. Cavalli-Sforza. 1996. Microsat: A computer program for calculating various statistics on microsatellite allele data . Leland Stanford Junior University, Palo Alto.

Mitani, J. C., K. L. Hunley, and M. E. Murdoch. 1999. Geographic variation in the calls of wild chimpanzees: a reassessment. *American Journal of Primatology* 47: 133-151.

Mitani, M. 1991a. Niche overlap and polyspecific associations among sympatric Cercopithecids in the Campo Animal Reserve, southwestern Cameroon. *Primates* 32: 137-151.

Mitani, M. 1991b. A note on the present situation of the primate fauna found from south-eastern Cameroon to northern Congo. *Primates* 31: 625-634.

Miyamoto, M. M. and J. Cracraft, eds. 1991. *Phylogenetic Analysis of DNA Sequences*. New York: Oxford University Press.

Moore, J. 1992. Savanna Chimpanzees. In *Topics in Primatology, Vol I: Human Origins*, ed. T. Nishida, W. McGrew, W. C. Marler, P. Pickford, and F. B. M. d. Waal:99-118. Tokyo: University of Tokyo Press.

Moore, J. 1996. Savanna chimpanzees, referential models and the last common ancestor. In *Great Ape Societies*, ed. W. C. McGrew, L. F. Marchant, and T. Nishida:275-292. Cambridge: Cambridge University Press.

Moore, S. S., L. L. Sargeant, T. J. King, J. S. Mattick, M. Georges, and D. J. S. Hetzel. 1991. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous primers pairs in closely related species. *Genomics* 10: 654-660.

Moreau, R. E. 1969. Climatic changes and the distribution of forest vegetation in Africa. *Journal of the Zoological Society of London* 158: 39-61.

Morin, P. A. 1992. Population Genetics of Chimpanzees. Ph.D., University of California, San Diego.

Morin, P. A. 1993. Reproductive strategies in chimpanzees. *Yearbook of Physical Anthropology* 36: 179-212.

Morin, P. A., S. Kanthaswamy, and D. G. Smith. 1997. Simple sequence repeat (SSR) polymorphisms for colony management and population genetics in rhesus macaques (*Macaca mulatta*). *American Journal of Primatology* 42: 199-213.

Morin, P. A., P. Mahboubi, S. Wedel, and J. Rogers. 1998. Rapid screening and comparison of human microsatellite markers in baboons: allele size is conserved, but allele number is not. *Genomics* 53: 12-20.

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

Morin, P. A., J. Wallis, J. J. Mooore, R. Chakroborty, and D. S. Woodruff. 1993. Non-invasive sampling and DNA amplification for paternity exclusion, community structure and phylogeography in wild chimpanzees. *Primates* 34: 347-356.

Morin, P. A., J. Wallis, J. J. Moore, and D. S. Woodruff. 1994b. Paternity exclusion in a community of wild chimpanzees. *Molecular Ecology* 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*, ed. R. D. Martin, A. F. Dixon, and E. J. Wickings:63-81. Basel: Karger.

Moritz, C. 1994. Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology* 3: 401-411.

Morrone, J. J. and J. V. Crisci. 1995. Historical Biogeography: Introduction to methods. *Annual Review of Ecology and Systematics* 26: 373-401.

Nei, M. 1977. F-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics* 41: 225-233.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from small numbers of individuals. *Genetics* 89: 583-590.

Nei, M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nei, M., F. Tajima, and Y. Tateno. 1983. Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution* 19: 153-170.

Nei, M. and N. Takahata. 1993. Effective population size, genetic diversity, and coalescence time in subdivided populations. *Journal of Molecular Evolution* 37: 240-244.

Nichol, J. E. 1999. Geomorphological evidence and Pleistocene refugia in Africa. *The Geographical Journal* 165: 79-89.

Nishida, T. 1979. The social structure of chimpanzees of the Mahale Mountains. In *The Great Apes.*, ed. D. A. Hamburg and E. R. McCown:73-121. Menlo Park: Benjamin/Cummings.

Nomiyama, H., M. Fukada, S. Wakesugi, T. Tsuzuki, and K. Shimada. 1985. Molecular structures of mitochondrial-like-DNA sequences in the human nuclear genome. *Nucleic Acids Research* 13: 1649-1658.

O'Brien, S. J. 1994. A role for molecular genetics in biological conservation. *Proc Natl Acad Sci U S A* 91: 5748-5755.

O'Brien, S. J. and E. Mayr. 1991. Bureaucratic mischief: Recognizing endangered species and subspecies. *Science* 251: 1187-1188.

Oates, J. F. 1988. The distribution of *Cercopithecus* monkeys in West African forests. In *A Primate Radiation: Evolutionary Biology of the African Guenons*, ed. A. Gautier-Hion, F. Boulière, J. P. Gautier, and J. Kingdon:79-103. Cambridge: Cambridge University Press.

Oates, J. F., IUCN/SSC Primate Specialist Group., and International Union for Conservation of Nature and Natural Resources. Species Survival Commission. 1996. *African primates : status survey and conservation action plan.* IUCN/SSC action plans for the conservation of biological diversity. Gland, Switzerland: IUCN.

Oates, J. F. (1999). *Myth and Reality in the Rain Forest: How Conservation Strategies are Failing in West Africa.* Berkeley: University of California Press.

Oates, J. F., P. A. Anadu, E. L. Gadsby, and J. L. Werre. 1992. Sclater's Guenon. *National Geographic Research and Exploration* 8: 476-491.

Oates, J. F., T. T. Struhsaker and G. H. Whitesides 1997. Extinction faces Ghana's Red Colobus monkey and other locally endemic species. *Primate Conservation* 17: 138-144.

Oates, J. F., D. White, E. L. Gadsby, and P. O. Bisong. 1990. Cross River National Park (Okwangwo Division), Feasibility Study: Appendix 1. U.K.: WWF.

Oden, N. L. 1984. Assessing the significance of a spatial autocorrelagram. *Geographical Analysis* 15: 1-16.

Olsen, G., J. Matsuda, R. Hagstrom, and R. Overbeek. 1994. FastDNAml: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Computer Applications in the Biosciences* 10: 41-48.

Oppo, D. W., J. F. McManus, and J. L. Cullen. 1998. Abrupt climate events 500,000 to 340,000 years ago: evidence from subpolar north atlantic sediments. *Science* 279: 1335-1338.

Osmaston, H. A. 1989. Glaciers, glaciations and equilibrium line altitudes on Kilimanjaro. In *Quaternary and Environmental Research on East African Mountains*, ed. W. C. Mahaney:7-30. Rotterdam: Balkema.

Palmer, M. R., P. N. Pearson, and S. J. Cobb. 1998. Reconstructing past ocean pH-depth profiles. *Science* 282: 1468-1471.

Pamilo, P. and M. Nei. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution* 5: 568-583.

Parsons, T. J., D. S. Muniec, K. Sullivan, N. Woodyatt, R. Alliston-Greiner, M. R. Wilson, D. L. Berry, K. A. Holland, V. W. Weeden, P. Gill, and M. M. Holland. 1997. A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics* 15: 363-368.

Pastorini, J., M. R. Forstner, R. D. Martin, and D. J. Melnick. 1998. A reexamination of the phylogenetic position of *Callimico* (primates) incorporating new mitochondrial DNA sequence data. *Journal of Molecular Evolution* 47: 32-41.

Paszek, A. A., G. H. Flickinger, L. Fontanesi, C. W. Beattie, G. A. Rohrer, L. Alexander, and L. B. Schook. 1999. Evaluating evolutionary divergence with microsatellites. *Molecular Evolution* 46: 121-126.

Pesole, G., C. Gissi, A. De Chirico, and C. Saccone. 1999. Nucleotide substitution rate of mammalian mitochondrial genomes. *Molecular Evolution* 48: 427-434.

Pesole, G., E. Sbisa, G. Preparata, and C. Saccone. 1992. The evolution of the mitochondrial D-loop region and the origin of modem man. *Molecular Biology and Evolution* 9: 587-598.

Plumptre, A. J. and V. Reynolds. 1997. Nesting behaviour of chimpanzees: Implications for censuses. *International Journal of Primatology* 18: 475-485.

Prescott, J., W. A. Rapley, and M. M. Joseph. 1996. Status and conservation of chimpanzee and gorilla in Cameroon. *Primate Conservation* 14-15: 7-12.

Primmer, C. R. and H. Ellegren. 1998. Patterns of molecular evolution in avian microsatellites. *Molecular Biology and Evolution* 15: 997-1008.

Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution* 16: 1791-1798.

Pusey, A. 1979. Intercommunity transfer of chimpanzees in Gombe National Park. In *The Great Apes*, ed. D. A. Hamburg and E. R. McCown:465-479. Menlo Park: Benjamin/Cummings.

Queller, D. C., J. E. Strassman, and C. R. Hughes. 1993. Microsatellites and kinship. *Trends in Ecology and Evolution* 8: 285-288.

Ramin, J. 1976. *Le Périple d'Hannon*. BAR supplementary series ; 3. Oxford: British Archaeological Reports.

Reynolds, V. and G. Luscombe. 1970. On the existence of currently recognized subspecies of chimpanzee (*Pan troglodytes*). *Folia Primatologica* 14: 129-138.

Rial, J. A. 1999. Pacemaking the ice ages by frequency modulation of Earth's orbital eccentricity. *Science* 285: 564-568.

Rithidech, K. N., J. J. Dunn, and C. R. Gordon. 1997. Combining multiplex and touchdown PCR to screen murine microsatellite polymorphisms. *Biotechniques* 23: 36, 40, 42, 44.

Robbins, C. B. 1978. The Dahomey Gap--A reevaluation of its significance as a faunal barrier to West African high forest mammals. *Bulletin of the Carnegie Museum of Natural History* 6: 168-174.

Rogers, A. R. 1995. Genetic evidence for a Pleistocene population explosion. *Evolution* 49: 608-615.

Rogers, A. R., A. E. Fraley, M. J. Bamshad, W. S. Watkins, and L. B. Jorde. 1996. Mitochondrial mismatch analysis is insensitive to the mutational process. *Molecular Biology and Evolution* 13: 895-902.

Rogers, A. R. and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9: 552-569.

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

Rothschild, W. 1905. Notes on anthropoid apes. *Proceedings of the Zoological Society of London for 1904* : 413-440.

Rothschild, W. 1906. Further notes on anthropoid apes. *Proceedings of the Zoological Society of London* : 465-468.

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. *Molecular Biology and Evolution* 9: 575-586.

Rubinsztein, D. C., W. Amos, J. Leggo, S. Goodburn, S. Jain, S. H. Li, R. L. Margolis, C. A. Ross, and M. A. Ferguson-Smith. 1995. Microsatellite evolution--evidence for directionality and variation in rate between species. *Nature Genetics* 10: 337-343.

Rubinsztein, D. C., J. Leggo, and W. Amos. 1995. Microsatellites evolve more rapidly in humans than in chimpanzees. *Genomics* 30: 610-612.

Ruvolo, M. 1994. Molecular evolutionary processes and conflicting gene trees: the hominoid case. *American Journal of Physical Anthropology* 94: 89-113.

Ruvolo, M. 1996. A new approach to studying modern human origins: hypothesis testing with coalescence time distributions. *Molecular Phylogenetics and Evolution* 5: 202-219.

Ruvolo, M. 1997. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. *Molecular Biology and Evolution* 14: 248-265.

Ruvolo, M., D. Pan, S. Zehr, T. Goldberg, T. Disotell, and M. von Dornum. 1994. Gene trees and hominoid phylogeny. *Proceedings of the National Academy of Sciences USA* 91: 8890-8904.

Ruvolo, M., S. Zehr, M. von Dornum, D. Pan, B. Chang, and J. Lin. 1993. Mitochondrial COII sequences and modern human origins. *Molecular Biology and Evolution* 10: 1115-1135..

Saccone, C., G. Pesole, and E. Sbisa. 1991. The main regulatory region of mammalian mitochondrial DNA: Structure-function model and evolutionary pattern. *Journal of Molecular Evolution* 33: 83-91.

Sachs, J. P. and S. J. Lehman. 1999. Subtropical North Atlantic Temperatures 60,000 to 30,000 Years Ago. *Science* 286: 756-759.

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

Sayer, J., C. S. Harcourt, and N. M. Collins, eds. 1992. *Conservation Atlas of Tropical Forests, Africa*. New York: MacMillian.

Scherer, R. P., A. Aldahan, S. Tulaczyk, G. Possnert, H. Engelhardt, and B. Kamb. 1998. Pleistocene collapse of the West Antarctic Ice Sheet. *Science* 281: 82-85.

Schiøtz, A. 1967. The treefrogs (Rhacophoridae) of West Africa. *Zoologica Musei Hauniensis* 25: 1-346.

Schneider, S. and L. Excoffier. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites. Application to human mitochondrial DNA. *Genetics* 152: 1079-1089.

Schneider, S., J. M. Kueffer, D. Roessli, and L. Excoffier. 1997. Arlequin: A software for population genetic data analysis 1.1. Genetics and Biometry Laboratory, University of Geneva, Geneva.

Schug, M. D., C. M. Hutter, K. A. Wetterstrand, M. S. Gaudette, T. F. C. Mackay, and C. F. Aquardro. 1998. The mutation rates of di-, tri and tetranucleotide repeats in *Drosophila melanogater*. *Molecular Biology and Evolution* 15: 1751-1760.

Schwarz, E. 1934. On the local races of chimpanzee. *Annals & Magazine of Natural History, London* 13: 576-583.

Shea, B. T. and H. J. Coolidge. 1988. Craniometric differentiation and systematics in the genus *Pan*. *Journal of Human Evolution* 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. *Human Biology* 66: 761-775.

Shriver, M. D., L. Jin, R. Chakroborty, and E. Boerwinkle. 1993. VNTR allele frequency distributions under the stepwise mutation model. *Genetics* 134: 983-993.

Shriver, M. D., L. Jin, R. E. Ferrell, and R. Deka. 1997. Microsatellite data support an early population expansion in Africa. *Genome Research* 7: 586-591.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science* 236: 787-792.

Slatkin, M. 1989a. Detecting small amounts of gene flow from phylogenies of alleles. *Genetics* 121: 609-612.

Slatkin, M. 1989b. Population structure and evolutionary progress. *Genome* 31: 196-202.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. *Genetics Research* 58: 167-175.

Slatkin, M. 1993. Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* 47: 264-279.

Slatkin, M. 1994a. Gene flow and population structure. In *Ecological Genetics*, ed. L. A. Real:3-17. Princeton: Princeton University Press.

Slatkin, M. 1994b. Cladistic analysis of DNA sequence data from subdivided populations. In *Ecological Genetics*, ed. L. A. Real:18-34. Princeton: Princeton University Press.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* : 457-462.

Slatkin, M. and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially 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.

Slatkin, M. and W. P. Maddison. 1990. Detecting isolation by distance using phylogenies of genes. *Genetics* 126: 249-260.

Smith. 1989. *Evolutionary Genetics*. New York: Oxford University Press.

Smouse, P. E. and R. Peakall. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82: 561-573.

Sneath, P. H. A. and R. R. Sokal. 1973. *Numerical Taxonomy*. San Franciso: W. H. Freeman.

Sokal, R. R. and N. L. Oden. 1978a. Spatial autocorrelation in biology 1. methodology. *Biological Journal of the Linnean Society* 10: 199-228.

Sokal, R. R. and N. L. Oden. 1978b. Spatial autocorrelation in in biology 2: some biological implications and four applications of evolutionary and ecological interest. *Biological Journal of the Linnean Society* 10: 229-249.

Soulé, M. E., ed. 1986. *Conservation Biology: The Science of Scarcity and Diversity*. Sunderland, MA: Sinauer Associates, Inc.

Soulé, M. E., ed. 1987. *Viable Populations for Conservation*. Cambridge: Cambridge University Press.

Tajima, M. and N. Takazaki. 1994. Estimation of evolutionary distance for reconstructing molecular phylogenetic trees. *Molecular Biology and Evolution* 11: 278-286.

Tajima, N. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-95.

Takahata, N. and S. R. Palumbi. 1985. Extranuclear differentiation and gene flow in the finite island model. *Genetics* 109: 441-457.

Takahata, N. and M. Slatkin. 1984. Mitochondrial gene flow. *Proceedings of the National Academy Sciences U S A* 81: 1764-1767.

Talbot, M. R. and G. Delibrias. 1977. Holocene variations in the level of Lake Bosumtwi, Ghana. *Science* 268: 722-724.

Talbot, M. R., D. A. Livingstone, P. G. Palmer, J. Maley, J. M. Melack, G. Delibrias, and S. Gulliksen. 1984. Preliminary results from sediment cores for Lake Bosumtwi, Ghana. *Paleoecology of Africa* 16: 173-192.

Tamura, K. and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10: 512-526.

Taylor, J. S., J. S. Sanny, and F. Breden. 1999. Microsatellite allele size homoplasy in the guppy (*Poecilia reticulata*). *Journal of Molecular Evolution* 48: 245-247.

Teleki, G. 1989. Population status of wild chimpanzees (*Pan troglodytes*) and threats to survival. In *Understanding Chimpanzees*, ed. P. G. Heltne and L. A. Marquardt:312-353. Cambridge, MA: Harvard University Press.

Templeton, A. R. 1991. Genetics and conservation biology. In *Species Conservtion: A Population-Biological Approach*, ed. A. Seitz and V. Loeschcke:15-29. Basel: Birkhauser-Verlag.

Templeton, A. R. 1993. The "Eve" hypothesis: A genetic critique and reanalysis. *American Anthropologist* 95: 51-72.

Templeton, A. R. 1994. Biodiversity at the molecular genetic level: Experiences from disparate macroorganisms. *Philosophical Transactions of the Royal Society of London B* 345: 59-64.

Thompson, L. G., M. E. Davis, E. Mosley-Thompson, T. A. Sowers, K. A. Henderson, V. S. Zagorodnov, P. Lin, V. N. Mikhalenko, R. K. Campen, J. F. Bolzan, J. Cole-Dai, and B. Francou. 1998. A 25,000-year tropical climate history from Bolivian ice cores. *Science* 282: 1858-1864.

Valdes, A. M., M. Slatkin, and N. B. Freimer. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* 133: 737-749.

Viard, F., P. Franck, M. P. Dubois, A. Estoup, and P. Jarne. 1998. Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *Journal of Molecular Evolution* 47: 42-51.

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.

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. *Proceedings of the National Academy of Sciences, USA* 88: 8720-8724.

Watanabe, Y., T. Sakazume, K. Kurosaki, H. Oota, K. Washio-Watanabe, and S. Ueda. 1997. No directionality of short tandem repeat evolution at the HPRT locus in catarrhine primates. *Folia Primatologica* 68: 350-353.

Weber, B. S. and C. Wong. 1993. Mutation model of human short tandem repeats. *Human Molecular Genetics* 2: 1123-1128.

Weir, B. 1996a. *Genetic Data Analysis*. Sunderland, MA: Sinaeur Associates, Inc.

Weir, B. 1996b. Intraspecific differentiation. In *Molecular Systematics*, ed. D. M. Hillis, C. Moritz, and B. K. Mable:385-406. Sunderland, MA: Sinaeur Associates, Inc.

Weir, B. S. and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

Western, D. and M. C. Pearl, eds. 1989. *Conservation for the Twenty-First Century*. New York: Oxford University Press.

White, F. 1983. The Vegetation of Africa. France: UNESCO.

Whiten, A., J. Goodall, W. C. McGrew, T. Nishida, V. Reynolds, Y. Sugiyama, C. E. Tutin, R. W. Wrangham, and C. Boesch. 1999. Cultures in chimpanzees. *Nature* 399: 682-685.

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.

Williamson, L. and L. Usongo. 1996. Survey of gorillas (*Gorilla gorilla*) and chimpanzees (*Pan troglodytes*) in the Reserve de Fauna du Dja, Cameroon. *African Primates* 2: 67-72.

Wilson, D. E. and D. M. Reeder. 1993. *Mammal Species of the World: A Taxonomic and Geographic Reference*. Washington: Smithsonian Institution Press.

Wise, C. A., M. Sraml, D. C. Rubinsztein, and S. Easteal. 1997. Comparative nuclear and mitochondrial genome diversity in humans and chimpanzees. *Molecular Biology and Evolution* 14: 707-716.

Woodruff, D. S. 1990. Genetics and the conservation of animals in fragmented habitats. In *International Conference on Tropical Biodiversity, "In Harmony with Nature"*:258-272. Kuala Lumpur, Malaysia: Malay Nature Society.

Woodruff, D. S. 1993. Non-invasive genotyping of nonhuman primates. *Primates* 34: 333-346.

Wright, S. 1951. The genetical structure of populations. *Annals of Eugenics* 1: 323-334.

Wright, S. 1969. *Evolution and the Genetics of Populations 2. The Theory of Gene Frequencies.* Chicago: University of Chicago Press.

Yang, Z. and A. D. Yoder. 1999. Estimation of the transition/transversion rate bias and species sampling. *Journal of Molecular Evolution* 48: 274-283.

Zardoya, R., D. M. Vollmer, C. Craddock, T. Streelman, S. Karl, and A. Meyer. 1996. Evolutionary conservation of microsatellite flanking regions and their use in the resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *Proceedings of the Royal Society of London B* 263: 1589-1598.

Zharkikh, A. 1994. Estimation of evolutionary distances between nucleotide sequences. *Journal of Molecular Evolution* 39: 315-329.

Zhivotovsky, L. A., M. W. Feldman, and S. A. Grishechkin. 1997. Biased mutations and microsatellite variation. *Molecular Biology and Evolution* 14: 926-933.

Zischler, H., H. Geisert, and J. Castersana. 1998. A hominoid-specific nuclear insertion of the mitochondrial d-loop: Implications for reconstructing ancestral mitochondrial sequences. *Molecular Biology and Evolution* 15: 463-469