

mtDNA Sequence Diversity of Orangutans from the Islands of Borneo and Sumatra

C. Cam Muir,* B.M.F. Galdikas,** Andrew T. Beckenbach

Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, B.C. V5A 1S6, Canada

Received: 3 February 1999 / Accepted: 2 August 2000

Abstract. A comparison of mitochondrial DNA sequences was undertaken for two genes among orangutans from Borneo and Sumatra. The distribution of haplotypes among 42 individuals for NADH dehydrogenase subunit 3 and 39 individuals for cytochrome B was used to infer population structure. The haplotypes among all Bornean orangutans form a cluster of closely related individuals for both genes, with two distinct haplotypes occupying different regions of the island. Sumatran haplotypes fall into three distinct, and highly diverged, groups. Strikingly, one of the Sumatran haplotypes shares sequence identity with the most widespread Bornean haplotype. This haplotype distribution is considered in the context of the highly complex geological history for the area around the Malay Archipelago. Alternating periods of geographic isolation and reunion, resulting from glacially induced land bridge formation, presented substantial opportunity for population dispersal between periodically isolated demes. We present a paleodispersal model that is consistent with genetic, geological, paleoecological, and fossil data. The disparity of sequences present in orangutan populations argues against a simple Sumatra–Borneo dichotomy. Our evidence, and that of others, suggests that the Sumatran population alone contains the remnants of at least three separate lineages.

Key words: *Pongo* — Orangutan — Borneo — Sumatra — mtDNA — Sundaland

Introduction

The orangutan (*Pongo pygmaeus*) is, for a variety of reasons, the least understood of the five great ape species: *Pan troglodytes*, *P. paniscus*, *Gorilla gorilla*, *Pongo pygmaeus*, and *Homo sapiens*. Presently, the orangutan is listed as a CITES Appendix 1 Endangered Species (most endangered) and its distribution is now limited to the islands of Borneo and Sumatra. The history of the orangutan lineage is a complicated one, characterized by great dispersal over time. The common ancestor of great apes is believed to be of African origin (Moyà Solà and Köhler 1993). The first splinters from this group, *Dryopithecus* and then *Sivepithecus*, migrated out of Africa more than 12 Myr ago (Andrews and Cronin 1982; Kappelman et al. 1991; Moyà Solà and Köhler 1993). The ancestor of the orangutan made its way into Asia, leaving a relatively rich fossil record of *Pongo* species throughout South Asia, well up into China, and including Borneo, Sumatra, and Java (Drawhorn 1994). There is no verifiable evidence that orangutans have survived outside of Borneo and Sumatra since the last glacial retreat.

Contemporary orangutan populations have been designated separate subspecies according to their island of origin, and some authors have recently argued that the populations should be elevated to distinct species (Zhi et al. 1996; Xu and Arnason 1996b; but see Muir et al. 1998a). This taxonomy has been based on morphology (size, hair color, beard, size of cheek pads) and genetics (allozymes, nuclear RFLPs, mtDNA sequence, and chromosomal inversions), which seem to have correlations with the island of origin. However, widespread problems of subspecific classification in zoos have resulted from ambiguities in “definitive” morphological characters and

* Current address: Department of Biology, University of Hawaii at Hilo, Hilo, HI 96720, USA

** Department of Archaeology, Simon Fraser University, Burnaby, B.C. V5A 1S6, Canada

Correspondence to: A.T. Beckenbach; e-mail: abeck@haldane.mbb.sfu.ca

there are alternative interpretations of the genetic data. For example, considerable importance has been placed on the diagnostic power of orangutan karyotypes. An inversion identified on chromosome 2 appears to be correlated with island of origin (Seuanez et al. 1979). Since it is generally assumed that fixed inversion differences do not arise frequently, this observation, together with the apparent absence of heterozygotes in natural populations, is taken as evidence of ancient bifurcation within the species. Other chromosome rearrangements have been discovered, however, including a polymorphic inversion on chromosome 9 and deletions on chromosomes 14 and 22 (deBoer and Seuanez 1982). In the captive community, there are many Sumatran–Bornean crosses that are fully viable, and second-generation offspring of these crosses are fully fertile. The authors who reported these characters noted a Sumatran orangutan that is heterozygous for the chromosome 2 inversion (Seuanez et al. 1979). Furthermore, 6 of 58 orangutans in a different study lacked the expected karyotype for their island of origin (Ryder and Chemnick 1993). These authors explained away these anomalies as zoo clerical errors.

Sample sizes have been a serious problem with many genetic studies (Muir et al. 1998a). Part of the difficulty lies with the problems associated with studying endangered species: availability of samples. True random sampling is not feasible with endangered species, so we must analyze all available samples. This approach can be particularly problematic in charismatic organisms that are the subject of illegal poaching for the exotic pet trade. Many orangutans in zoos—those most readily available to researchers—were rescued from the pet trade and are of unknown or ill-defined origin.

In this paper we analyze sequence data from two mitochondrial genes, NADH dehydrogenase subunit 3 (*nad3*) and cytochrome B (*cob*), for 39–42 individuals from Borneo and Sumatra. These animals are a combination of wild-caught and captive orangutans. The main purpose is to determine the extent of sequence variation in the mitochondrial genomes within and between the two islands. Networks are presented to illustrate the phylogenetic relationships of the genes sampled. We interpret the results in the context of historical evidence, based on the fossil record, the geological history of the Malay Archipelago, and the paleoecology of Sundaland (the union of Borneo, Sumatra, Java, Peninsular Malaysia, Thailand, Vietnam, Cambodia, and the Sunda shelf by lowered sea level).

Materials and Methods

Samples

The analysis of population structure and historical changes to the overall range, and dispersal within the range, requires tissue samples from

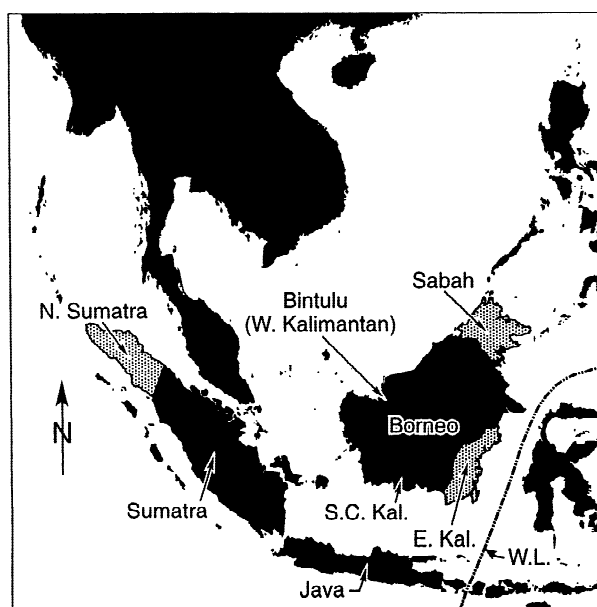


Fig. 1. Sample locales of Borneo and Sumatra orangutan populations. Contemporary populations of Sumatran orangutans are confined to the northern region of the island. The location of Wallace's Line (W.L.), which separates the Asian and Australian faunas, is also indicated on the map.

individuals across the present geographic range of the species. For samples to be most useful, their geographic origins must be well documented. There are a number of problems associated with adequate sampling of endangered species, and orangutans in particular. These problems include the need for noninvasive sampling and techniques to utilize small and forensic samples to minimize the insult of sampling. Comprehensive sampling of orangutan populations is also complicated by political boundaries found within the orangutan range.

There are a large number of captive orangutans in zoos and rehabilitation centers around the world. Unfortunately the records of the precise birth place some of these individuals are unavailable since many of these orangutans were confiscated from poachers. Wild orangutans are very difficult to sample since they are difficult to track, very shy, and arboreal. Samples used in this study are from three main sources.

Samples from Camp Leakey. Blood samples from 17 individuals were obtained by B. Galdikas from orangutans being rehabilitated at Camp Leakey in Tanjung Puting Kalimantan, Borneo (near the tip of the arrow, S.C. Kal., in Fig. 1). These orangutans are ex-captives and individuals confiscated from poachers. The exact origin of each of these orangutans is not known, although all are believed to have originated from Borneo. Heparinized and nonheparinized blood samples were kept at -70°C until used.

Samples from Captive Orangutans. Six samples were obtained from zoos across North America. These samples are a combination of blood and hair samples. There are no records of the exact origins of these orangutans except on which of the islands, Borneo or Sumatra, they originated. All of the GenBank sequences included in this study are from zoo-housed captive orangutans. Two of the Sumatran samples obtained for this study came from the Calgary zoo: Bella [International Studbook Number (ISN) 1980] and Mais II (ISN 1932). Ruby (ISN 2528), housed at the Miami Parrot Jungle and Gardens, is a Bornean/Sumatran cross. Her mother, Tasha (ISN 1760), is also a cross. Tasha's mother, Tammy (ISN 667), was a Sumatran (L Perkins, personal com-

munication) and so Ruby has a Sumatran maternal lineage. Abigail (ISN 525) and Jaura (ISN 2014) are both from Sumatra and both are housed at the Metropolitan Toronto Zoo. Two Bornean orangutans, Kelly (ISN 1793) and Doc (ISN 2009), are housed at the Houston Zoological Gardens.

Borneo Collections. In the summer of 1995 C. Muir went to Sabah and Sarawak to augment the sample set with tissues from orangutans from broader and better-documented origins. Twenty-four of the samples collected on this field trip were used in this study. Orangutans sampled as a result of being confiscated from poachers from Lahad Datu, Telupid, Kinabatangan, and Bahil-Garam Sandikan (all in Sabah) were kindly made available through a cooperative association with Mahedi Andau, Director of the Wildlife Department of Sabah, and Dr. Edwin Bossi, Chief Veterinarian at Sepilok. All of these orangutans were being rehabilitated at the Sepilok Orangutan Rehabilitation Center near Sandikan. Many were translocated from various palm oil plantations to the Tabin Wildlife Reserve. Translocated orangutans are identified alpha-numerically (except Tim = Timura Plantation) as follows: *Ynn*, Tunku Suan Lanka; *Xnn*, Abedon; *Hnn*, Hing Lee; *Wnn*, EAC; *Vnn*, Pahang 2. Additional samples were kindly made available through Sapuan Amaad, Director of Forestry and Wildlife in Sarawak, and Dr. Manabu Onuma, Chief Veterinarian at the Semingoh Rehabilitation Centre. These individuals were confiscated from poachers at Lubok Ntu, near Kuching, and Bintulu (West Kalimantan). Hair samples were plucked close to the skin to ensure that follicles were included. Hairs were kept in plastic bags and either frozen or placed in 70% EtOH. There did not appear to be any difference in the quantity or quality of DNA extracted from hairs stored in these two ways.

DNA Preparation

The method used for DNA extraction depended on the sample type. Blood samples were subject to Iso-Quick (Orca Research Inc.) extractions with minor revisions to the manufacturer's suggested protocol (Muir et al. 1995). Chelex 100 was used to extract DNA from hair follicles (Walsh et al. 1991). Because of limited quantities of some samples, it was not feasible to complete replicate extractions and sequencing from every individual for both genes. The sequence for both genes is available for 30 of the individuals.

General PCR and Sequencing Methods

Polymerase Chain Reaction (PCR). Whole genomic extracts were used for PCR reactions. In general, less than 50 ng of DNA was used per reaction, but individual extracts were not quantified due to the limited quantity of available samples. PCR reaction mixtures included 1 mM MgCl₂, *Taq* buffer (50 mM KCl, 10 mM Tris-HCl, 0.1% Triton X-100), 0.2 mM primer, 1 U *Taq* polymerase (Bio/Can). PCR was performed under the following conditions: 35 cycles of melting at 95°C for 1 min, annealing at 60°C for 1 min, and extending at 72°C for 1 min. Following the 35 cycles, a 5-min extension was used. Primers used for both PCR and sequencing of *nad3* are as follows: L9415, 5' CCATCTACTGATGAGGGTCTTAC; and H9975, 5' ATTAGGTGTGAGCGGTAGAC. *Cob* was amplified and sequenced using the following primers: ECB1, 5' CACGAAACAGGATCAAATAACCC; and ECB2, 5' ATTTTCAGTTACAAGGCTGGCG.

Sequencing. ³³P-labeled dideoxy terminator sequencing (Amersham) was used for direct manual sequencing from Prep-a-gene (Bio Rad), or "freeze and squeeze" (Tautz and Renz 1983), purified PCR products following manufacturers' suggested protocols (cycle sequencing reaction over 40 cycles of 90°C for 30 s, 55°C for 30 s, 72°C for 30 s). Sequence reaction products were separated on 6% urea poly-

acrylamide gels in glycerol-tolerant (taurine) buffer. Unique haplotypes encountered in this study are available from GenBank under the following accession numbers: AF273440–AF273445 (*nad3*) and AF273446–AF273450 (*cob*).

Analysis. Sequence alignment was done manually with ESEE3s (Cabot and Beckenbach 1989). Networks were drawn by hand. Maximum-likelihood and -parsimony analysis was carried out on concatenated *nad3* and *cob* sequences using Phylip [version 3.57c (Felsenstein 1994)]. Neighbor-joining, using Jukes–Cantor corrected distances, was conducted using MEGA [version 1.0 (Kumar et al. 1993)]. Homologous regions extracted from the following published complete mtDNA sequences were included as outgroups: chimpanzee [*Pan troglodytes* (Horai et al. 1995); GenBank accession D38113], bonobo [pygmy chimpanzee, *Pan paniscus* (Horai et al. 1995); D38116], gibbon [*Hyllobates lar* (Arnason et al. 1996); X99256], gorilla [*Gorilla gorilla* (Xu and Arnason 1996a); X93347], and human [*Homo sapiens* (Anderson et al. 1981); J01415.1].

Results

Complete sequences of the mitochondrial *nad3*, comprising 345 bp, were determined for 40 individuals, including 36 from Borneo and 4 from Sumatra. These sequences were aligned and compared with *nad3* from the complete mitochondrial sequences published by Horai et al. (1992) from a Bornean orangutan and by Xu and Arnason (1996b) for an individual from Sumatra. In addition, we determined partial sequences, 496 bp, from mitochondrial *cob* for 34 individuals, including 5 from Sumatra, and compared them with the comparable region from the two published complete sequences, two partial sequences determined by Xu and Arnason (1996b), and a partial sequence published by Collura and Stewart (1995). Of these the sequence for both genes is available for 30 individuals, including the two published complete sequences.

A total of nine haplotypes was observed among the 42 individuals compared for *nad3*, each differing by one or more nucleotide substitutions. Among the 39 individuals compared for *cob* sequences, a total of eight haplotypes was encountered. The haplotypes are indicated for each individual in Table 1. Haplotype variants for *nad3* are indicated by numbers; those for *cob*, by letters. Haplotypes 1–4 and A–D were encountered only on Sumatra. Haplotype 5 and E were found on both islands. Haplotypes 6–9 and F–H were observed only from Borneo. The sole or predominant island of origin is also indicated: Sum or Bor. When taken together, the 30 orangutans for which the sequence was available for both genes included 12 distinct haplotypes, differing in either *nad3*, *cob*, or both.

Nad3 haplotype *Bor-9* was observed in 14 individuals, but 13 of those were from a series of plantations in Sabah, in northern Borneo. The only other *Bor-9* haplotype was from West Kalimantan, to the south of Sabah. The only widespread haplotype, *Bor-5*, was found in West Kalimantan, South Central Kalimantan, and one individual from Sumatra.

Table 1. Names, origins and haplotype designations of samples used in this study^a

Origin	Name	<i>nad3</i>	<i>cob</i>	Composite
Sumatra	X97707 ^b	<i>Sum-1</i>	<i>Sum-A</i>	<i>Sum-1A</i>
	Abigail	<i>Sum-2</i>	<i>Sum-B</i>	<i>Sum-2B</i>
	Mais II	<i>Sum-3</i>	<i>Sum-C</i>	<i>Sum-3C</i>
	Bella	<i>Sum-4</i>	<i>Sum-D</i>	<i>Sum-4D</i>
	Ruby	<i>Bor-5</i>	<i>Bor-E</i>	<i>Bor-5E</i>
	Gambar	—	<i>Bor-E</i>	—
	Jaura	—	<i>Bor-E</i>	—
	Ini ^c	—	<i>Sum-B</i>	—
	"Borneo"	D38115 ^d	<i>Bor-5</i>	<i>Bor-F</i>
Davida		<i>Bor-5</i>	<i>Bor-E</i>	<i>Bor-5E</i>
Diane		<i>Bor-6</i>	<i>Bor-E</i>	<i>Bor-6E</i>
Doc		<i>Bor-7</i>	<i>Bor-G</i>	<i>Bor-7G</i>
Kelly		<i>Bor-5</i>	<i>Bor-E</i>	<i>Bor-5E</i>
CLO1239		<i>Bor-5</i>	—	—
CLO48		<i>Bor-8</i>	—	—
Anna ^b		—	<i>Bor-E</i>	—
Dennis ^b		—	<i>Bor-E</i>	—
South Central Kalimantan	Brook	<i>Bor-5</i>	<i>Bor-E</i>	<i>Bor-5E</i>
	Hobler Lily	<i>Bor-5</i>	<i>Bor-E</i>	<i>Bor-5E</i>
	Julie	<i>Bor-5</i>	<i>Bor-E</i>	<i>Bor-5E</i>
	Mark	<i>Bor-6</i>	<i>Bor-E</i>	<i>Bor-6E</i>
	Stan	<i>Bor-5</i>	<i>Bor-E</i>	<i>Bor-5E</i>
	Rosemary	<i>Bor-5</i>	—	—
West Kalimantan	Bebeta	<i>Bor-9</i>	<i>Bor-E</i>	<i>Bor-9E</i>
	Herbie	<i>Bor-5</i>	<i>Bor-E</i>	<i>Bor-5E</i>
	Maggie	<i>Bor-5</i>	<i>Bor-H</i>	<i>Bor-5H</i>
	Patti	<i>Bor-5</i>	<i>Bor-E</i>	<i>Bor-5E</i>
	Roger	<i>Bor-5</i>	<i>Bor-E</i>	<i>Bor-5E</i>
	Siswi	<i>Bor-7</i>	<i>Bor-E</i>	<i>Bor-7E</i>
	Supinah	<i>Bor-5</i>	<i>Bor-E</i>	<i>Bor-5E</i>
	Lemot	<i>Bor-7</i>	—	—
	"Baboon"	<i>Bor-7</i>	—	—
	East Kalimantan	Apollo Bob	<i>Bor-8</i>	<i>Bor-E</i>
Mellie		<i>Bor-8</i>	<i>Bor-E</i>	<i>Bor-8E</i>
Sabah	Tim	<i>Bor-9</i>	<i>Bor-E</i>	<i>Bor-9E</i>
	Gensusuli	<i>Bor-9</i>	<i>Bor-E</i>	<i>Bor-9E</i>
	Hing Lee	<i>Bor-9</i>	<i>Bor-E</i>	<i>Bor-9E</i>
	Kim Long	<i>Bor-9</i>	<i>Bor-E</i>	<i>Bor-9E</i>
	Semenduh	<i>Bor-9</i>	—	—
	Ah Fong	—	<i>Bor-G</i>	—
	H2	<i>Bor-9</i>	—	—
	H5	—	<i>Bor-E</i>	—
	V03	<i>Bor-9</i>	—	—
	V04	<i>Bor-9</i>	<i>Bor-E</i>	<i>Bor-9E</i>
	W01	—	<i>Bor-G</i>	—
	W02	<i>Bor-9</i>	—	—
	W03	<i>Bor-9</i>	—	—
	W05	<i>Bor-9</i>	<i>Bor-E</i>	<i>Bor-9E</i>
	W07	<i>Bor-9</i>	—	—
X01	<i>Bor-9</i>	—	—	
Y01	—	<i>Bor-E</i>	—	

^a Individuals labeled "Borneo" are from unknown or unidentified localities on that island. All localities listed below "Borneo" are specific regions of Borneo. Dashes indicate that the sequence was not determined

^b Xu and Arnason (1996b)

^c Collura and Stewart (1995)

^d Horai et al. (1992)

Three sequences of *cob* were encountered more than once. Abigail is identical over the 496 bp determined in this study to Ini [GenBank accession U38274 (Collura and Stewart 1995)]. Both are Sumatran. Their *cob* haplotype is indicated as *Sum-B*. Two individuals from Sabah, and one whose origin is identified only as "Borneo," shared the *Bor-G* haplotype. There were 28 individuals sampled from every region in Borneo and from Sumatra that shared the *Bor-E* haplotype for *cob*. These individuals include Ruby, who was zoo-bred but whose records indicate an unbroken Sumatran maternal lineage, Jaura, of Sumatran origin, and Gambar, who was caught in the wild in Sumatra.

The maximum divergence observed within Borneo was only 0.4%, but within Sumatra sequence divergence of over 9% was observed for the two genes combined (Table 2). These within-Sumatra divergences are as great as the most distant pairwise comparisons between Borneo and Sumatra sequences (Table 2). These results suggest that Borneo sequences form a very tight group, with most haplotypes differing by only one or a few nucleotide substitutions, whereas Sumatra includes an extraordinary genetic diversity.

Geographic differentiation of *nad3* sequences is evident within Borneo. All 13 individuals sampled from Sabah, in northern Borneo, shared a single haplotype, *Bor-9*, differing from the widespread *Bor-5* haplotype at two sites. Among nine individuals from West Kalimantan, in the region bordering Kalimantan and Sarawak, three haplotypes were observed, but only one individual carried the *Bor-9* haplotype predominant to the north.

Phylogenetic Relationships Among Haplotypes

Phylogenetic illustration of relatedness between genetic types is typically given in the form of a gene tree. These gene trees are of two basic formats: (1) a simple bifurcating tree that constrains genetic types to occupy only branch tips and (2) a network joining related haplotypes which allows them to occupy internal nodes. In the case of intraspecific comparison, it is likely that "ancestral" alleles are represented in contemporary populations. In this case, constraining ancestral alleles to occupy only branch tips in a bifurcating tree may result in a misleading illustration of relationships (Crandall and Templeton 1993).

Figures 2 and 3 give networks showing the inferred relationships among the observed haplotypes for *nad3* and *cob*. They show very similar results for the two genes. In both networks, there are three quite distinct groups. One group includes all of the orangutans sampled from Borneo. The second group is represented in our sample by a single individual, Bella (*Sum4*, *Sum-D*). The third group is a relatively loose assemblage of "Sumatran" orangutans, including the sequence published by Xu and Arnason (1996b).

Table 2. Pairwise comparisons of the *nad3* and *cob* sequences determined in this study^a

	Sumatra haplotype				Borneo haplotype							
	<i>Sum-1A</i>	<i>Sum-2B¹</i>	<i>Sum-3C</i>	<i>Sum-4D</i>	<i>Bor-5E</i>	<i>Bor-5F</i>	<i>Bor-6E</i>	<i>Bor-8E</i>	<i>Bor-5H</i>	<i>Bor-9E</i>	<i>Bor-7E</i>	<i>Bor-7G</i>
<i>Sum-1A</i>	—	7/11	4/14	25/43	26/36	26/37	27/36	27/36	26/37	26/36	27/36	27/35
<i>Sum-2B¹</i>	2.25	—	7/13	26/46	26/39	26/40	27/39	27/39	26/40	24/39	27/39	27/38
<i>Sum-3C</i>	2.17	2.50	—	23/49	26/40	26/41	27/40	27/40	26/41	26/40	27/40	27/39
<i>Sum-4D</i>	8.56	9.44	9.09	—	20/25	20/26	21/25	20/25	20/24	20/25	21/25	21/26
<i>Bor-5E</i>	7.76	8.47	8.29	5.55	—	0/1	1/0	1/0	0/1	2/0	1/0	1/1
<i>Bor-5F</i>	7.89	8.60	8.42	5.68	0.12	—	1/1	1/1	0/2	2/1	1/1	1/2
<i>Bor-6E</i>	7.89	8.60	8.42	5.68	0.12	0.24	—	2/0	1/1	3/0	2/0	2/1
<i>Bor-8E</i>	7.89	8.60	8.42	5.55	0.12	0.24	0.24	—	1/1	3/0	2/0	2/1
<i>Bor-5H</i>	7.89	8.60	8.42	5.42	0.12	0.24	0.24	0.24	—	2/1	1/1	1/2
<i>Bor-9E</i>	7.76	8.19	8.29	5.55	0.24	0.36	0.36	0.36	0.36	—	3/0	3/1
<i>Bor-7E</i>	7.89	8.60	8.42	5.68	0.12	0.24	0.24	0.24	0.24	0.36	—	0/1
<i>Bor-7G</i>	7.76	8.47	8.29	5.81	0.24	0.36	0.36	0.36	0.36	0.48	0.12	—

^a Upper triangle: number of differences, *nad3/cob*. Lower triangle: Jukes–Cantor distances for both genes combined. Total sites compared: 345 for *nad3*; 496 for *cob*

^b Only 316 sites analyzed for *nad3*

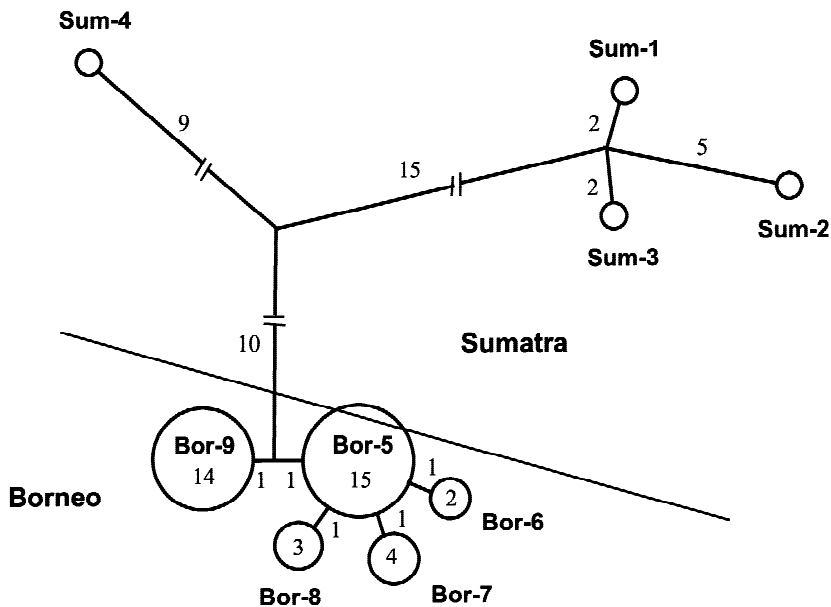


Fig. 2. Network of *nad3* haplotypes. Distances are indicated as the number of differences on each branch. For haplotypes that were observed more than once, the number of individuals carrying each haplotype is indicated in the circles.

Because of the high degree of divergence observed among some of the orangutan haplotypes, it is useful to construct a bifurcating tree, as well. Maximum likelihood, with or without the assumption of a molecular clock, maximum parsimony, and neighbor-joining gave identical topologies, after collapsing the Borneo haplotypes. Figure 4 shows the neighbor-joining tree based on concatenated sequences of all observed haplotypes. The tree includes the homologous regions extracted from published complete mitochondrial sequences of the other species of great apes and is rooted by the gibbon sequence. Bootstrap sampling strongly supports the existence of three distinct groups of orangutan haplotypes, with *Sum-4D* somewhat closer to the Bornean sequence assemblage than to the other Sumatran group.

The Possibility of Nuclear Pseudogenes

The high level of intraspecific divergence reported here (up to 9%), based on PCR amplification and direct sequencing, raises the possibility that nonfunctional nuclear inserts are contributing to the variation. Indeed, Collura and Stewart (1995) have reported nuclear inserts of *cob*, including the region sequenced in this study, from a Sumatran orangutan and from other hominoids. While there are no reports of nuclear *nad3* inserts, the existence of such sequences cannot be disproved. There are several lines of evidence suggesting that our results reflect variation in the functional mitochondrial genes. First, comparisons of known nuclear *cob* pseudogene sequences with the sequences determined in this study re-

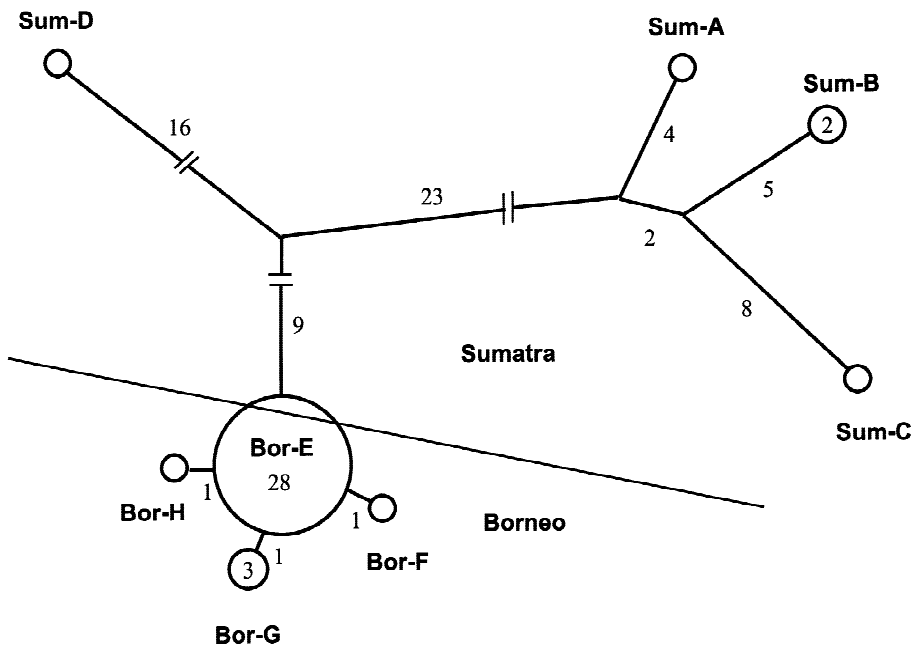


Fig. 3. Network of *cob* haplotypes. Distances are indicated as the number of differences on each branch.

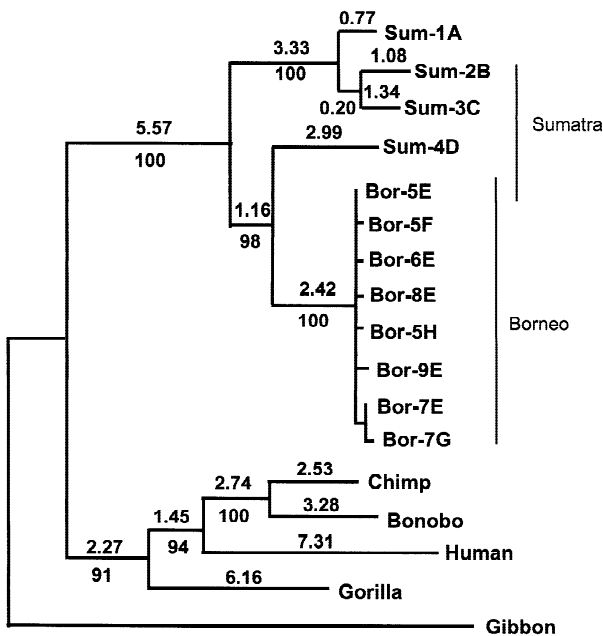


Fig. 4. Neighbor-joining tree of concatenated haplotypes. Distances were corrected using the Jukes–Cantor correction. Numbers above the branches are branch length; bootstrap values (of 100) are given below the branches.

veal a number of indels, resulting in frameshifts, as well as a large number of substitutions in the nuclear pseudogenes. All of the sequences analyzed here have open reading frames, as expected for functional mitochondrial genes. Second, two of the most divergent sequences, those published by Horai et al. [(1992); *Bor-5F*] and Xu and Arnason [(1996b); *Sum-1A*], are based on complete mitochondrial sequences and clearly represent divergences in the functional mitochondrial genes. Third, the two genes studied here, which are widely separated in the

genome, give comparable results for all individuals. Bella (*Sum-4D*), for instance, differs by over 7% in both genes from the published Sumatran sequence (*Sum-1A*) and 5–6% from the published Borneo sequence (*Bor-5F*) (Table 2).

Discussion

Sequence comparisons of *nad3* sequences from 42 individuals (38 from Borneo and 4 from Sumatra) and partial sequences of *cob* from 39 individuals (32 Borneo and 7 Sumatra) have revealed an extraordinary degree of variation in these two genes. Sequences from Borneo form a tight cluster, while those from Sumatra fall into three very distinct lineages. We believe that the divergence evident in both genes within Sumatra must have resulted from an extended period of isolation of at least three groups. We believe that their presence together on Sumatra is a relatively recent event. The diversity observed in Sumatra is in stark contrast to the tight clustering of sequences observed from Borneo. The difference in diversity on the two islands exists despite the fact that present-day populations in Sumatra are much smaller than those on Borneo, and are restricted to a much smaller geographic area, in northern Sumatra. We propose a paleogeographic explanation for these observations, based on the geological history of the region, paleoecology, and evidence from the morphology of fossil and extant orangutans.

Glacial Effects on Local Geography

During major glacial advances precipitation is accumulated in advancing glaciers, to the extent that the sea

levels drop significantly. Several such glacial epochs have occurred over the past million years and have resulted in sea levels dropping as much as 120 m (Van der Kaars and Dam 1997). The effects of the changing sea level are most dramatic where the coastal shelf is shallow, as is the case in the area west of Wallace's line known as the Sunda Shelf. With a drop in the sea level of as little as 30 m, land bridge formation begins between Sumatra and Peninsular Malaysia and also between Borneo and Sumatra. In fact the Sunda Shelf was emergent almost continuously from 500 to 120 kya (Batchelor 1979; Drawhorn 1994). By the time the glacial advance maximized, all of Borneo, Sumatra, and Java, some of the Philippines, and the Asian mainland to Vietnam had become a single landmass known as Sundaland. Most recently these land bridges persisted from about 60,000 years ago until as recently as 10,000 years, or half of the last 100,000 years.

Geology

Approximately 74,000 years ago the Toba Volcano, in northern Sumatra, produced the second largest volcanic explosion for which there is evidence on this planet. The 10 billion metric tons of ash thrown into the atmosphere resulted in a global drop in temperatures. The eruption of Toba has, in fact, been attributed with bringing on the Wisconsin Glacial Epoch, although there is some debate about the magnitude of its influence on initiating the Wisconsin Epoch (Rampino and Self 1992; Kerr 1996). There seems little doubt that a disturbance of this magnitude would have a catastrophic effect on ecosystems, especially the larger herbivorous inhabitants, for some distance. Since that time the area surrounding the Toba Caldera has been reclaimed by the orangutan.

Paleoecology

Sequestering of water into the ice sheets during glacial maxima would have drastically lowered sea levels, connecting the islands of Borneo, Java, and Sumatra to the South Asian mainland. Fossil pollen analysis from piston cores taken from the Sunda shelf indicates the presence of tropical savanna and diptocarp woodland forests on the land bridge for much of their existence (Adams 1993; Van der Kaars and Dam 1997), suggesting that the habitat was suitable for orangutan populations. During glacial maxima lowered temperatures and vegetation zones (Drawhorn 1994; Van der Kaars and Dam 1997) may have created the impetus for dispersal from highlands.

The DNA evidence indicates that Borneo populations constitute a close assemblage of haplotypes. We hypothesize that the orangutan population distribution in Borneo has been stable since the colonization of that region and that distinct haplotypes (*Bor-5* and *Bor-9*) are of recent origin. Most of the population in Sumatra was

probably extirpated as a result of the Toba explosion. Present populations in northern Sumatra appear to be the result of recolonizations from three or four sources: Borneo, mainland Asia, southern Sumatra, and Java. The maps shown in Fig. 5 represent a time series over two major glacial epochs and were developed using bathymetric data from the United States Geological Survey database as described by Muir et al. (1998b). In the first frame, representing a period at least 200,000 to 300,000 years ago, the sea levels are shown to be similar to those today. The ancestral distribution of orangutans, through South Asia but prior to colonization of Borneo, Sumatra, and Java, is indicated. As the sea level dropped to approximately 40 m lower than that today (frame two), land bridges began to appear between the mainland and the islands. The arrow indicates one of many possible dispersal routes leading to the colonization of Sumatra (Drawhorn 1994). As the sea level dropped to -125 m at the glacial maximum, the islands of the archipelago west of Wallace's Line (W.L.) became a single land mass joined to the mainland. The curved arrow in the third frame indicates proposed colonization of the entire land mass by orangutans. Wallace's Line (Figs. 1 and 5) separates the Asian and Australian faunas and corresponds to a deep ocean strait separating Bali from Lombok (south of Java) and Borneo from Celebes (Darlington 1957, pp. 462-472).

As sea levels rose with the retreating glacier, populations on Borneo (Z), Sumatra (Y), and the mainland (X) would have become isolated from each other as shown in the fourth and fifth frames. These populations would have remained isolated until at least the next glacial period. During the subsequent glacial period(s) the opportunity existed for populations, which may have become genetically distinct, to disperse from Borneo and the mainland into Sumatra along routes proposed by Drawhorn (1994), indicated in frames 6, 7, and 8. As indicated in frame 9 the mainland and Java populations of orangutans have become extinct, while Sumatra supports genotypes which are highly diverged, one of which shares identity with Bornean orangutans.

Fossil Evidence of Diversity

There is substantial morphometric diversity among fossil orangutans from across South Asia (Drawhorn 1994). Even contemporary morphological diversity is great and reflects the genetic diversity we found. In fact, Drawhorn (1994) points out morphometric similarities between fossils found on Borneo and contemporary orangutans found in Sumatra. It had been suggested previously that two or more populations exist in Sumatra (deBoer 1982). Drawhorn also points out that the assertion that Sumatran and Bornean populations have been isolated as distinct groups for more than a million years (as suggested by Zhi et al. 1996; Xu and Arnason 1996b; Janczewski et al.

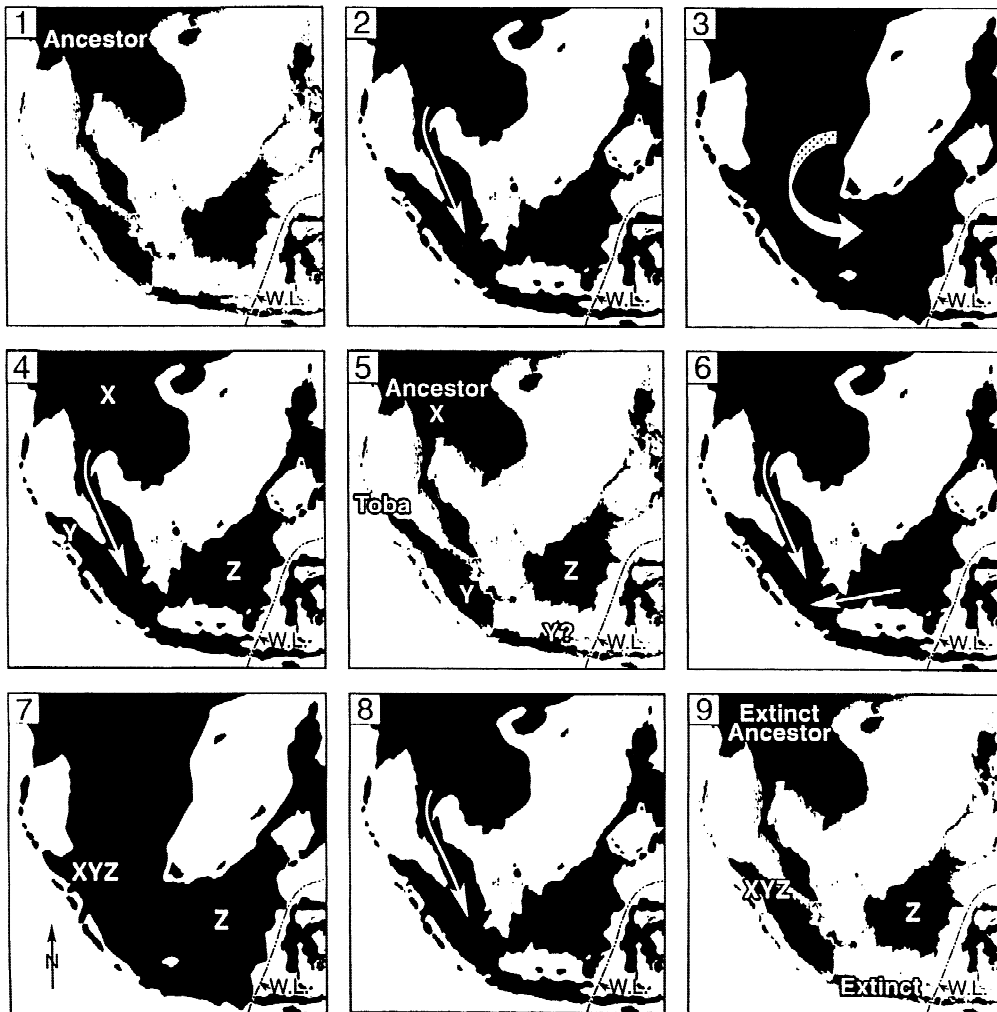


Fig. 5. Paleodispersal predicted based on haplotype distribution and availability of route. Maps show the effect of falling sea level on the presence of land bridges. *Arrows* indicate possible dispersal routes. The

map series is from Muir et al. (1998b), represents two full glacial advance and retreat cycles, and runs left to right, top to bottom.

1990) is not supported by the fossil record, even though that record is well documented. Drawhorn points out that the derived character states that distinguish these populations are not evident until the recent fossil record.

In summary, the fossil evidence shows that 80,000 years ago, and for a long time afterward, there were orangutan populations in much of the South Asian mainland, Sumatra, Borneo, and Java. When the Toba volcano exploded, about 74,000 years ago, it probably destroyed orangutan populations in most of Sumatra—at least in the north. Shortly after this, the Wisconsin Glacial Epoch began its advance. As the glacier sequestered water, extensive land bridge formation occurred. Lasting for about 50,000 years, the land bridges probably sustained robust tropical woodland ecosystems (Hantoro 1997), while much of the present lowland forests were replaced by open swamp, dominated by grasses and sedges (Van der Kaars and Dam 1997). Being highly opportunistic, the orangutan is very likely to have taken advantage of open niches presented on the land bridge

and reclaimed area surrounding the Toba caldera. The mtDNA sequence divergence evident in the remnant populations on Sumatra appear to reflect multiple, diverse origins.

The Species Question

The high level of divergence (about 8%) between a complete mtDNA sequence from a Sumatran orangutan and that from a Bornean orangutan has been used as evidence of species level divergence between the populations on those two islands (Xu and Arnason 1996b). Although, as noted in the Introduction, there are other genetic distinctions, as well as subtle morphological differences between individuals from Sumatra and Borneo (for a summary see Zhi et al. 1996), the mtDNA sequence divergence appears to be the most compelling evidence yet. We have argued elsewhere (Muir et al. 1998a) that comparison of two mtDNA sequences, even complete

ones, is not sufficient evidence to define species. Results presented here suggest that orangutan population structure is more complex than a simple dichotomy of Sumatra and Borneo populations. The Borneo populations, sampled widely from the island, form a compact collection of mtDNA haplotypes. Sumatra, in contrast, appears to include at least three very distinct lineages, including one of the Borneo lineages.

Other authors have encountered "Borneo" karyotypes or morphotypes in Sumatran orangutans but have dismissed the observations as misclassification (Ryder and Chemnick 1993). We observed "Borneo" haplotypes from several individuals that could be traced to Sumatra: Ruby, of Sumatran maternal lineage, Gambar, and Jaura. While one could argue that they either are misclassified or result from recent introductions by humans, no simple Sumatra/Borneo dichotomy can explain Bella. Bella's *nad3* sequence differed from the "Sumatra" haplotype at 25 sites (*Sum-4* versus *Sum-1*: 7.2%); her *cob* haplotype differed at 43 sites (*Sum-D* versus *Sum-A*: 8.7%). This divergence is comparable to the difference between "Borneo" and "Sumatra" complete mtDNA sequences (Horai et al. 1992; Xu and Arnason 1996b). She also differs from the Borneo cluster of haplotypes by 5.8 and 5.0% at these two genes (Table 2). This result suggests that her mitochondrial lineage has been isolated from both "Borneo" and "Sumatra" lineages for an extended period of time.

We believe that geographic changes over the last two or more glacial/interglacial cycles can explain the high genetic diversity that is currently present in Sumatra. Geographically, Sumatra occupies a central position among Borneo, Java, Sumatra, and the South Asian mainland (Fig. 1). These islands have undergone repeated episodes of long-term isolation, followed by extended periods of land bridge connections. The intriguing possibility that the distinct haplotypes found in Bella, or the "Sumatran" group of haplotypes, are descendants of now extinct populations from Java or the South Asian mainland is testable if DNA can be obtained from subfossil remains from those populations.

Acknowledgments. We gratefully acknowledge the help of Patrick Andau, Sapaun Ahmad, Edwin Bossi, Manabu Onuma, Misuari Elis, Joe Fred, Chip Fletcher, Jane and Gordon Tribble, Hugues Faure, Ian Singleton, Sander Van der Kaars, Peter Kershaw, Lloyd Burckle, and Wyss Yim for assistance in obtaining samples for this study. Al Arndt, Charlene Mayes, Ian Tan, Karen Beckenbach, Russ Watkins, Thé Ha, Mike Vankoeveringe, Rebecca Cann, and Mark Stoneking assisted in various aspects of the analysis. Several anonymous reviewers made numerous suggestions leading to substantial improvements in the manuscript. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to A.T.B.

References

Adams JM (1993) Towards a better vegetation system for global mapping and monitoring. Background paper for the UNEP–HEM Charlottesville Workshop, Jan

- Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465
- Andrews P, Cronin JE (1982) The relationship of *Sivapithecus* and *Ramapithecus* and the evolution of the orang-utan. *Nature* 297:541–545
- Arnason U, Gullberg A, Xu X (1996) A complete mitochondrial DNA molecule of the white-handed gibbon, *Hylobates lar*, and comparison among individual mitochondrial genes of all hominoid genera. *Hereditas* 124:185–189
- Batchelor BC (1979) Discontinuously rising late Cenozoic sea levels with specific reference to Sundaland, Southeast Asia. *Geol Mijnbouw* 58:1–20
- Cabot EL, Beckenbach AT (1989) Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. *Comp Appl Biosci* 5:233–234
- Collura RV, Stewart C (1995) Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature* 378:485–489
- Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* 134:959–969
- Darlington PJ (1957) *Zoogeography*. Wiley, New York
- deBoer LEM (1982) Genetics and conservation of the orangutan. In: deBoer LEM (ed) *The orangutan. Its biology and conservation*. Dr. W. Junk, The Hague, pp 39–61
- deBoer LEM, Seunanez HN (1982) The conservation of the species. In: deBoer LEM (ed) *The orangutan. Its biology and conservation*. Dr. W. Junk, The Hague, pp 135–171
- Drawhorn GM (1994) The systematics and paleodemography of fossil orangutans (genus *Pongo*), PhD thesis. University of California, Davis
- Felsenstein J (1994) PHYLIP (phylogenetic inference package), version 3.5c. Distributed by the author. University of Washington, Seattle
- Hantoro WS (1997) Quaternary sea level variations in the Pacific-Indian Ocean gateways: response and impact. *Quaternary Int* 37:73–80
- Horai S, Satta Y, Hayasaka K, Kondo R, Inoue T, Ishida T, Hayashi S, Takahata N (1992) Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J Mol Evol* 35:32–43
- Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci USA* 92:532–536
- Janczewski DN, Goldman D, O'Brien SJ (1990) Molecular genetic divergence of orangutan (*Pongo pygmaeus*) subspecies based on isozyme and two-dimensional gel electrophoresis. *J Hered* 81:375–387
- Kappelman J, Kelley J, Philbeam D, Sheikh KA, Ward S, Anwar M, Barry JC, Brown B, Hake P, Johnson NM, Raza SM, Shah S (1991) The earliest occurrence of *Sivapithecus* from the middle Miocene Chinji Formation of Pakistan. *J Hum Evol* 21:61–73
- Kerr RA (1996) Volcano-ice age link discounted. *Science* 272:817
- Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis, version 1.0. Pennsylvania State University, University Park
- Moyà Solà S, Köhler M (1993) Recent discoveries of *Dryopithecus* shed new light on evolution of great apes. *Nature* 365:543–545
- Muir CC, Galdikas BMF, Beckenbach AT (1995) Genetic variability in orangutans. In: Nadler R, Galdikas BMF, Rosen N (eds) *The neglected ape*. Plenum Press, New York, pp 267–272
- Muir CC, Galdikas BMF, Beckenbach AT (1998a) Is there sufficient evidence to elevate the orangutan of Borneo and Sumatra to separate species? *J Mol Evol* 46:378–381
- Muir CC, Watkins RF, Ha T (1998b) Map simulations, based on bathometric, and glacial induced sea level change data, are used to

- test a putative paleodispersal hypothesis inferred from orangutan population genetics data. *assemblage* (1998 4th ed. <http://www.shef.ac.uk/~assem/4/>)
- Rampino MR, Self S (1992) Volcanic winter and accelerated glaciation following the Toba super-eruption. *Nature* 359:50–52
- Ryder OA, Chemnick LG (1993) Chromosomal and mitochondrial DNA variation in orangutans. *J Hered* 84:405–409
- Seuanez HN, Evans HJ, Martin DE, Fletcher J (1979) An inversion in chromosome 2 that distinguishes between Bornean and Sumatran orangutans. *Cytogenet Cell Genet* 23:137–140
- Tautz D, Renz M (1983) An optimized freeze-squeeze method for the recovery of DNA fragments from agarose gels. *Anal Biochem* 132:14–19
- Van der Kaars S, Dam R (1997) Vegetation and climate change in West-Java, Indonesia during the last 135,000 years. *Quaternary Int* 37:67–71
- Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for PCR based typing from forensic material. *Biotechniques* 10:506–513
- Xu X, Arnason U (1996a) A complete sequence of the mitochondrial genome of the western lowland gorilla. *Mol Biol Evol* 13:691–698
- Xu X, Arnason U (1996b) The mitochondrial DNA molecule of Sumatran and a molecular proposal for two (Bornean and Sumatran) species of orangutan. *J Mol Evol* 43:431–437
- Zhi L, Karesh WB, Kanczewski DN, Frazier-Taylor H, Sajuthi D, Gombek F, Andau M, Martenson JS, O'Brien SJ (1996) Genomic differentiation among natural populations of orang-utan (*Pongo pygmaeus*). *Curr Biol* 6:1326–1336