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

THE DESCENDANTS OF OUR ANCESTORS: INVESTIGATING POPULATION STRUCTURE OF THE ORANGUTAN (*PONGO PYGMAEUS*) USING DNA SEQUENCE AND PALEOMIGRATION MODELING

by

Cedric Campbell Muir
B.Sc. Simon Fraser University 1991

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in the
Department of Biological Sciences

SIMON FRASER UNIVERSITY
July 1998

Abstract

Up until 40 000 years ago, the orangutan (*Pongo pygmaeus*) ranged widely over South Asia. Sea level fluctuations which have resulted from numerous glacial advances have given rise to widespread landbridge formation between the islands of Borneo, Sumatra, and Java, with the South Asian mainland. Compilation of a sea level change series over a bathometric matrix was used to create a video which demonstrates changes in the land form resulting from glacial induced sea level fluctuations. Land bridges which existed provided possible migration routes between demes which presently share alleles. Thirteen mitochondrial and nuclear loci were sequenced and compared between up to 41 individuals from different sample areas. Phylogenetic analysis of these data indicate recent gene flow between some sample areas including presently distinct islands of Borneo and Sumatra. Rates of sequence evolution were compared across mitochondrial and nuclear loci including a (mt) gene/ (nuc) pseudo-gene pair. Coalescent theory, and a derivation from this theoretical framework, were engaged to infer which alleles are most like the ancestral state which contextualizes a paleomigratory hypothesis to explain contemporary allele distribution. Gene trees were constructed as cladograms which allow extant taxa to occupy internal nodes and are consistent with the paleomigratory model. Continued habitat destruction contributes, both directly and indirectly, to important declines in the overall numbers of orangutans as well as fragmenting remaining populations. Large clear-cut swaths may also limit gene flow between areas which have shared migrations.

Dedication

Always close to my heart, I dedicate this work to Tim and all of his species.

(Tim was the first wild orangutan that I met. He was trapped in the Timura plantation in Sabah Malaysia. Patrick, Edwin, Misuari, Elis, Joe Fred, and I moved Tim to the Tabin wildlife reserve in the summer of 1995.)

Acknowledgements: Andy, Biruté, Mike, Felix, Mark, Karen, Chuck, Ian, Al, Patrick, Edwin, Sapuan, Misuari, Elis, Joe Fred, Russ, Chip Fletcher, Jane, and Gordon Tribble, Hugues Faure, Sander VanderKaars, Peter Kershaw, Lloyd Burckle, Wyss Yim, Researchers in Quaternary Science mail list quaternary@morgan.ucs.mun.ca, Eliah, Franklin, Libbie

Aloha Na Holholona a me Pau Ola

(trans.: With Love for Animals and All Life)

Table of Contents

List of Appendices

List of Figures

List of Tables

List of Videos

Video 1: A simulation of the effects of changing sea levels on the geography surrounding Borneo , Sumatra, Java, and the South Asian mainland.

Video 2: A simulated "fly-by" around Sumatra, across the Malay peninsula, Borneo, and Java at present sea level.

Video 3: A simulation of proposed paleomigrations to explain contemporary population structure of the orangutan.

Videos are found at:

http:\\darwin.mbb.sfu.ca\imbb\beckenbach\cam.html
or on enclosed CD
An "avi" viewer is required to view videos.

Chapter 1

Introduction

Originally named *Simia satyrus* by Linneaus, the Orang Utan (orangutan, People of the forest, Mias, or *Pongo pygmaeus*) is the third closest living relative to human after the chimpanzee and gorilla (Ruvolo et al. 1994 Horai et al. 1992). Some time around twelve million years ago populations of an ancestor of living apes became separated. One branch is likely to have been ancestral to African Apes. The other branch migrated out of Africa and lead to the genus *Dryopithecus* which inhabited Europe and Asia (Martin and Andrews 1993). Subsequent bifurcations lead to *Lufengpithecus* and *Sivapithecus*, both of which are more similar to *Pongo* than to other Apes (Andrews and Cronin 1982, Martin and Andrews 1993, Solar and Kohler 1993). Fossils from both *Lufengpithecus* and *Sivapithecus* have been found in East Asia and predate the arrival of modern *Pongo* species. Fossil and sub-fossil remains of recent *Pongo* species have been found throughout South Asia (Andrews and Cronin 1982).

Presently, the orangutan occupies an almost entirely arboreal or canopy dwelling habitat. Some have argued whether the orangutan, or a sibling species, may have held a predominately terrestrial habit during the time prior to the Wisconsin era, 70 thousand years ago (kya)-present, in the South Asian mainland (Smith and Pilbeam 1980, Andrews and Cronin 1982, Galdikas 1981). This possibility raises the question whether arboreal habit is ancestral or recently derived. In either case, the orangutan occupied the Asian mainland until 40 kya or more recently (Drawhorn 1994). Perhaps coincidentally, this time frame roughly coincides with a period when Asia was probably experiencing an elevation in the population size of *Homo sapiens*. Present distribution of the orangutan is

limited to the islands of Sumatra and Borneo.

Fossil evidence places the orangutan in Java, Sumatra, and Borneo prior to the Last Glacial Maximum (LGM) (Smith and Pilbeam 1980). Orangutans must have arrived on these isolated islands via a land bridge that connected these three islands with the Asian mainland (Peltier 1994, Muir et al.. in review a and b). The extirpation of the orangutan from all but Borneo and Sumatra also followed vast ecological shifts that resulted from the second largest volcanic explosion known about 70 kya (Chesner et al. 1991). A global temperature drop of 3-5 C has been attributed to this event and some have argued that the Toba eruption precipitated the onset of the Wisconsin Glacial period (Rampino and Self 1992). Widespread ecological shifts preceding the glacial front may have provided an impetus for southward migration, or dispersal, of a variety of species.

Presently orangutan survival is endangered by various activities associated with contemporary human development, particularly logging and the establishment of palm oil plantations. There are probably fewer than 25,000 orangutans living in the wild today and to make matters worse, the effective population size is probably less than 8,000 given the extreamly long birth intervals, and the likelihood that only a small number of males are probably participating in reproduction (Galdikas 1984). Recent, out of control anthropogenic fires throughout Borneo and Sumatra have resulted, both directly and indirectly, in the death of many orangutans. A thriving, though illegal, exotic pet trade also threatens the survival of the orangutan. It has been estimated that seven to eight orangutans are killed, or otherwise die, for each baby orangutan that is successfully delivered to be a pet. Since the orangutan makes such a terrible choice for a pet they are often discarded early in their life. Given the precarious population size and slow

reproductive rate, contemporary demographic, genetic, and environmental/ habitat impacts may result in imbalances increasing susceptibility to extinction (Jones 1997).

The orangutan is an intelligent animal and is thought by some to think in a way very similar to humans. Orangutans, both captive and wild, have a high level of demonstrated problem solving ability. In captive settings, the orangutan is known as the Houdini of the zoo. Using a combination of brute strength, observation, trickery, and complex problem solving, orangutans have been known to break, or disassemble a cage, or trick the keeper into leaving the keys.

Working with an orangutan named Chantek, Lynn Miles has been communicating quite fluently in American Sign Language (ASL) with the ape for more than 20 years (Miles 1995). In the wild, orangutans are not seen to be extensive tool users. However, spontaneous, inventive, and copied tool use is common in captive orangutans. Captive orangutans are also more frequent communicators than their wild counter-parts. These differences do not indicate different abilities but more likely availability of time for introspection, retrospection, and inspection. In 1690 John Locke pointed out that liberation from the business of survival allows for reflection, consideration, and assimilation of knowledge (Locke 1690).

From an anthropological perspective, the orangutan has been described as semi-social (Galdikas 1985 a and b). Females associate more frequently than males though less frequently and for shorter duration than adolescents. Females rarely associate in a group for more than a day making them largely solitary but social (Galdikas 1984). Perhaps a well-dispersed population best suits their foraging and nest building requirements. Sub-adult males are more social than adult males. Males also occupy a

much larger range than females do. Females raise their young for a comparatively long time, staying with them until early adolescence. As a consequence, they have a birth interval up to 12 years (Galdikas and Wood 1990).

Most of the foraging time is spent eating fruit. The orangutan diet includes fruit, fungus, insects, honey, seeds, leaves, small vines, and bark. The change in their diet varys with season and habitat (Galdikas 1988). Males and females also seem to have different dietary composition (Galdikas 1982).

Orangutans serve an important role in their ecosystem. They are important seed dispersers as a consequence of their largely frugivore diet (Galdikas 1982). Eating terminal meristems releases lateral meristems from terminal dominance allowing for increased branch growth. Males are known to push down and dislodge jungle snags as often as one every 18 hours. This activity opens the canopy providing a necessary gap for diptocarp seedling growth. The result of this activity is the maintenance of age and species diversity of jungle flora and fauna.

The purpose of this research has been to infer population structure and migration dispersal history for the orangutan of Borneo and Sumatra. Molecular insight into population structure comes from analysis of allele distribution and frequency. Alleles are defined by one or more single nucleotide differences. Loci to be sequenced were chosen from both the mitochondrial and nuclear genomes. Animal mitochondrial genes are a good choice for intraspecific comparison because they have a high evolutionary rate comparred with nuclear genes of a comparable functional class. Mitochondrial genomes are inherited along a maternal lineage which, since it is haploid, is not subject to recombination though crossing over between the multiple genomes present within a given

mitochondrion is possible. Another advantage in using mitochondrial loci over nuclear loci comes into play with the study of endangered species for which only limited sample material is available. The effective concentration of the mitochondrial genome per cell is many times higher than the nuclear genome. The ribosomal cistron, which may be present in thousands of concatomeric copies, is a notable exception to the general pattern that the nuclear genes are present in only two copies, one paternally and the other maternally inherited. The mitochondrial genome, on the other hand, is present in multiple copies per mitochondrion since it replicates like bacterial DNA and its biosynthesis is not linked to the cell cycle. Furthermore each cell houses many mitochondria; metabolically active cells may have hundreds of mitochondria. Faced with the scenario of having to amplify DNA, by means of the Polymerase Chain Reaction (PCR) (Mullis and Fuloona 1987) from a single cell as an extreme example, you may still have a thousand copies of the mitochondrial locus of interest to act as a PCR template. The higher number of genomes per cell is a definite advantage where template DNA has been isolated from samples like hair or faeces that have a small number of genes from which target DNA can be amplified.

Sampling endangered species provides a number of challenges in terms of numbers of individuals which can be sampled and the amount of sample which is available. Wild orangutans are difficult to follow and get close enough to sample non-invasively and orangutan nests, which might provide hair samples are located high in the forest canopy.

Since the mitochondrial genome is inherited in a non-Mendelian, strictly maternal lineage, different range requirements of males and females prescribes the use of both

mitochondrial and nuclear loci in the choice of informative molecular markers. If there are sex specific differences in mating range, a paternally inherited gene is required to give insight into male population structure. In this work I chose to characterize a large number of genes which are maternally, bi-parentally, and paternally inherited. Mitochondrial loci include the tRNA genes for arginine (arg), lysine (lys), glycine (gly), and proline (pro), the protein coding genes for cytochrome oxidase subunit 2 (COII), nicotinamide adenine dinucleotide subunit 3 (ND3), and cytochrome b (Cyt b), and a 42 base spacer found between COII and lys. Nuclear loci included the fourth and sixth introns of the arylsulfatase gene, the third intron for lysozyme, and a nuclear pseudo-gene copy of the mitochondrial Cyt b gene. The Y-specific, paternally inherited, SRY locus (Whitfield et al. 1993) was also sequenced. Differences in the rate of evolution are examined for the two genomes. Among the markers studied are a mitochondrial gene and a nuclear copy pseudo-gene for cytochrome b.

Chapters 3, 5, and 8 of this thesis are modifications of manuscripts and a book chapter, which are in review or already published. Chapter 3 is modified from a manuscript (Muir et al. (a) in review) submitted to assemblage, and a chapter (Muir et al. (b)) in review for a book to be called Geoenvironmental Mapping: Theory, Method and Practice. Chapter 5 is modified from a manuscript (Muir et al.. (c) in review) submitted to Evolution. Chapter 8 contains a modified version of a manuscript published in JME (Muir et al. 1998)

Chapter 2
Land Bridge Predictions and 3-D Mapping Techniques

Introduction

This chapter describes a mapping technique that allows one to take available bathometric data matrices and convert them into a topology map over which water level can be modeled. There are a variety of applications for this technique including any task that involves predicting the effects of changing water levels to existing landscape. I will consider the application of this technique to the study of population genetics and the prediction of available routes for paleomigrations.

The study of population genetics and structure investigates, among other things, the genetic consequences of past and present dispersal, or gene flow, between isolated populations or demes. There are a number of factors that bear on the probability of dispersal between demes. Most fundamental of these influences is the accessability of the route: can the individuals cross the span in question? If the dispersal path is a long one, especially if it is likely to take longer than a single generation, the route must also be ecologically capable of sustaining the migrating population. Something must be known about the landscape in order to make predictions about possible gene flow between demes since a potential pathway may be rendered impassable by an intervening mountain ridge, major river or other barrier.

Past migrations affect present population structure and allele distribution. Dispersal options are influenced by a variety of factors that have not remained static over time. For example, islands, that are presently isolated by large bodies of water may have been historically connected by land bridges. Land bridges have appeared and retreated many times as a result of fluctuations in sea level resulting from major glacial advances.

During the Wisconsin Glacial Epoch and its retreat (~ 70 kya to present), global sea levels fell as much as 125 m lower than contemporary sea levels (Van der Kaars & Dam 1996, Chip Fletcher pers. comm., J. &. G. Tribble pers. comm.). A change in sea level of this magnitude had a dramatic impact on the size and shape of islands, peninsula, and continental coastal outlines.

The islands of Borneo, Sumatra, and Java, and their geographic relationship to the South Asian mainland provide a case in point. During the LGM (Last Glacial Maximum) these islands, in addition to the Malay Peninsula, Thailand, Vietnam, and Cambodia, all formed a much larger, single landmass. A number of options for range expansion or shifts were presented by these changes in limitations to terrestrial migrations (Fig. 1). As a result gene flow may have existed between populations that are presently isolated by great expanses of ocean.

Background

In order to predict routes that may have been available for dispersal one needs to be able to model intervening landscape between populations. Examination of any contemporary map reveals the barrier to movement of orangutans, or any terrestrially bound organism, between the islands of Borneo and Sumatra (i.e. the Karimata Straight). However, we know that sea levels have changed over time and that the landscape of ocean floors is not one great plain but is replete with mountains, valleys.

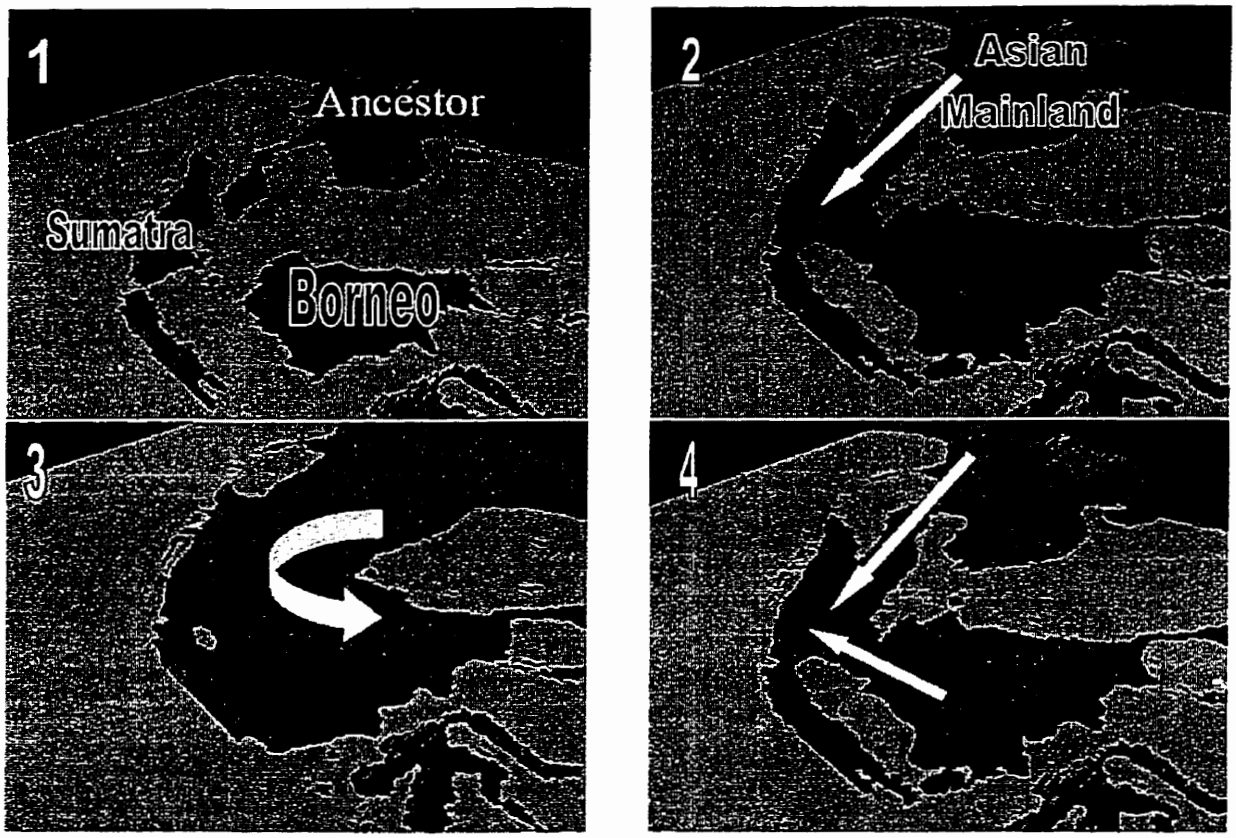


Figure 1: A series of maps which show the effect of glacial induced changes in sea level. Possible migratory routes shown with arrows.

It is not easy to predict the effect of changing sea level to above water landscape without a detailed topography both above and below water. Topographic information is organized in bathometric data sets that can be converted into two or three dimension representational maps.

Data Sources

Digital land height and bathometric data for the Borneo/Sumatra region was downloaded from the US National Geophysical Data Center web site, using the ETOPO5 database (Muir et al. 1998b). A 709 by 421 pixel data set was selected that included the region from latitude 27 degrees North to latitude 8 degrees South and 68 to 127 degrees

East Longitude. The data set downloaded from the web site consisted of a 1480003-byte space-delimited text file of elevations relative to sea level, ranging from -9394 meters to +4115 meters. The ETOPO5 database has a five arc-minute resolution, which corresponds to a width of approximately ten kilometers per data element, and thus provides a reasonably high-resolution view of the Borneo/Sumatra region. At present no higher resolution data set for land elevations in this region appears to exist in any public databases, with the exception of some small areas around Jakarta which are available in the GTOPO30 and GLOBE project databases. A newer database, the Measured and Estimated Seafloor Topography database (Muir et al. 1998b), may in fact contain better resolution data for below sea-level elevations in this region (on the order of four kilometers per data element), but was not utilized in this study. Estimates of sea level at a variety of points throughout the most recent glacial cycle were critical in attempting to simulate landform change. Much more limited amounts of data are available to document this in public databases on the Internet (Muir et al 1998b).

Examination and comparison of DNA sequences between individuals from throughout contemporary distribution allows one to establish which individuals are most closely related genetically. In the case that dispersal plays a role in population dynamics there may be individuals living in separate demes that are more closely related to each other than either is to other constituents of their respective demes. These relationships can be expressed with a map showing allele distribution. By showing allele distribution on a map one can infer how and where dispersal may have occurred between demes. Since past as well as present movement of populations play a role in allele distribution, a consideration of historic geography aids in establishing where dispersal was possible.

Methods for the preparation of two and three dimension maps and videos described here are available and can be found in Muir et al.. (in review a, and b).

Topographic maps can be illustrated two or three dimensionally (Figs. 1 and 2). Videos were also made using three dimensional map projections. The videos follow two different perspectives in viewing the effects of sea level change. In the first video (Video 1) a series of maps with serially decreasing followed by increasing sea levels gives a "satellite view" of the effect of changing sea level to the islands of Borneo, Sumatra, and Java. The second video (Video 2) shows a "helicopter view" of a single sea level. In this video one is given a land level "fly around" of prescribed circumference. The "helicopter view" gives a higher resolution view of landscape relief.

Mapping Relevance

The utility of this mapping technique to the study of population structure comes in attempting to understand allele distribution in the context of population expansion and contraction both past and present. For example, if one were interested in allele distribution among lakes, some of which have been connected as a result of increased water level, maps could be created that demonstrate the changing connectedness of lakes as a result of changing water levels. There is no difference, with respect to population genetics, between intermittent connections among bodies of water or land and their constituent populations.

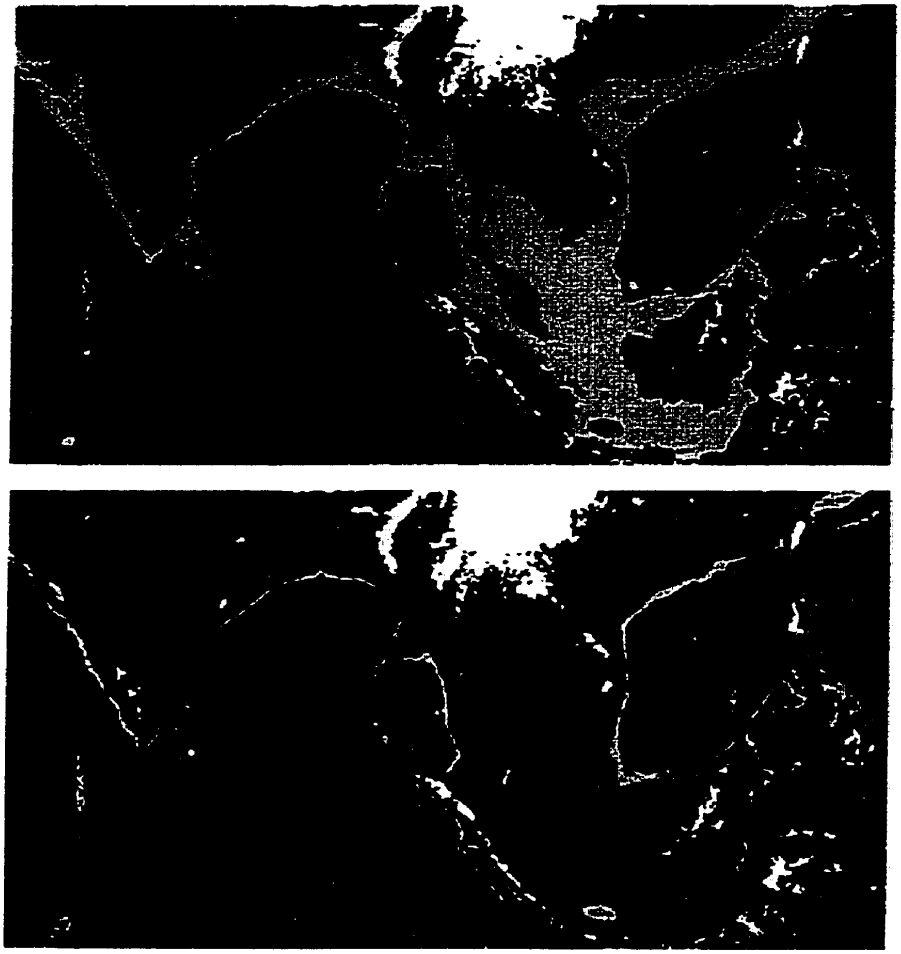


Figure 2a&b: South Asia map showing the effect of 120 m change in sea level on the formation of a land bridge joining the islands of Borneo, Java, and Sumatra, to the South Asian mainland. The top map shows present sea level while the bottom map shows sea level at glacial maximum. Light grey indicates shallow water.

Prior to the emergence of the Isthmus of Panama, gene flow was likely between the Gulf of Mexico and the Pacific Ocean but very little or no gene flow between Mexico and South America. The connection of North and South American landmasses resulted in blocking aquatic movements while opening up terrestrial dispersal possibilities

The historical patterns shown here will be used below to interpret results of genetic analysis (Chapters 5, and 6). The pattern of allele distribution makes sense by

allowing for the movement of populations across a land bridge between the two islands. Note in Fig. 1, which represents a full glacial cycle, that the land bridge is most persistent between the south tip of the Malay Peninsula and Sumatra, and between mid-Sumatra and the south west coast of Borneo. The land bridge connects Borneo at a point adjacent to the populations that share alleles with a Sumatran population. This coincident geographic location of shared alleles and land bridge connection suggests a recent dispersal of populations across the land bridge between the islands (Muir et al. in review a and b).

Future Prospects

The tremendous rate of increase in both processor speed and software complexity available to the average user is unlikely to abate in the near future. Simulation of ancient environments in near real-time is already within the grasp of high-end personal computers. The production of static images and short fly-through animations as demonstrated in this chapter is helpful to understanding the complex interactions of environment and population history that have resulted in the extant distribution of animals, but more complex simulations are already possible. A number of packages are available commercially, that allow the user to create a digital mesh of ancient landscapes from elevation and sea-level data, and to navigate through these landscapes at will. These programs could find some use in examining in detail specific features that might have affected population dynamics. Use of Auto-CAD software with landscape modeling programs will also provide an opportunity to input specific ecological data such as niche location, likely population density, species pairing and other predicted associations.

The limiting factor in the use of this technology is the availability and quality of

data that can be reliably used to make inferences about past landscapes. For instance, at the five arc-minute resolution of the data utilized in this chapter, local effects will be entirely unobservable. The increasing amount of 30 arc-second elevation data which is available in public databases will help to increase the resolution with which some aspects of ancient topography can be investigated, however local effects may well be obscured by error associated with estimates of sea level. Data of much higher resolution may slowly become available. Future prospects for utilizing landscape modeling in conjunction with population genetics, and ecological data provide only one avenue for the use of the mapping techniques described in this chapter.

Chapter 3

Methods

Samples

The analysis of population structure and historical changes to the overall range, and dispersal within the range, requires tissue samples from individuals across the present geographic range of the orangutan. In order that samples be most useful their geographic origin must be well documented. There are a number of problems associated with adequate sampling of endangered species and orangutans in particular. The importance of non-invasive sampling is obvious. Further development of PCR and DNA extraction techniques from small and forensic samples has been important in enabling the use of tissues that 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 centres around the world. Unfortunately the records of the precise birth place some of these individals are poor to non-existent 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. The samples used in this study are from three main sources (Appendix 1).

1) Samples from Biruté M.F. Galdikas. Twenty-one samples of blood and faecal samples were provided by B.M.F. Galdikas from orangutans being rehabilitated at Camp Leakey in Tanjung Puting Kalimantan. These orangutans are ex-captives and individuals confiscated from poachers. The exact origin of all of these orangutans is not known. Heparinized blood samples were taken by, or provided to, Biruté Galdikas at the Tanjung Puting rehabilitation center Camp Leakey in 1989 and were kept at -70 ° C until used.

2) Samples from captive orangutans. Six samples were obtained from zoos across North America (see Appendix 1). These samples are a combination of blood and hair samples. There are no records of the exact origins of these orangutans beyond from which of Borneo or Sumatra they hailed. Though not explicitly stated in each case, it seems that all GenBank samples included in this study are also from zoo housed captive orangutans. The number of samples from wild orangutans needs to increase. Two of the Sumatran samples, Bella, (International Studbook Number (ISN) 1980), and Mais II (ISN 1932) came from the Calgary Zoo. Ruby (ISN 2528), housed in 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 pers. comm.) and so Ruby has a Sumatran maternal lineage. Abigail (ISN 525) and Jaura (ISN 2014) are both from Sumatra and both are housed in the Metropolitan Toronto Zoo. Two Bornean orangutans, Kelly (ISN 1793) and Doc (ISN 2009), are housed in the Houston Zoological Gardens.

3) Field trip to Borneo. In the summer of 1995 I went to Sabah and Sarawak to augment my sample set with tissues from orangutans from a broader and more specific distribution. International field work, sampling endangered species, and in particular, an endangered primate, required extensive preparation. Appendix 2 provides a check list of preparations I found necessary to conduct this work. 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, Kinabalangan, and Bahil-Garam Sandikan were kindly made available through a cooperative association with Mahedi Andau the Director of the Wildlife Department of Sabah and Dr. Edwin Bossi

Chief Veternarian at Sepilok. All of these orangutans were being rehabilitated at the Sepilok Orangutan Rehabilitation Centre near Sandikan. Orangutans were translocated from various Palm Oil Plantations to the Tabin Wildlife Reserve. I received samples from many of these orangutans. Unfortunately, many of the early hair samples were taken with scissors and, therefore, did not include the necessary follicle. I was invited to participate in the liberation of orangutans from a number of Plantations in the area around Lahad Datu. Translocated Orangutans were identified alpha-numerically (except Tim = Timura Plantation) as follows: Y* = Tunku Suan Lanka, X* = Abedon, H* = Hing Lee, W* = EAC, V* = Pahang 2. Another set of samples were kindly made available through Sapuan Amaad, Director of Forestry and Wildlife in Sarawak, and Dr. Manabu Onuma, the Chief Veterinarian at the Semingoh Rehabilitation Centre. These individuals were confiscated from poachers at Lubok Ntu, near Kuching, and Bintulu. Hair samples were plucked close to the skin (orangutan hair is brittle and breaks easily without providing the necessary follicle). Hairs were kept in plastic bags and put into a freezer upon return. Some of the hair samples were kept 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 different ways. I collected faecal samples using a leaf to avoid contamination by human DNA.

DNA Preparation

I used my own DNA for a control sequence. My DNA was extracted from both blood and hair follicles.

The method used for DNA extraction depended on the sample type. Blood and

fecal samples were subject to Iso-Quick (Orca Research Inc.) extractions with minor revisions to manufacturer's suggested protocol (Muir et al. 1994). Chelex 100 was used to extract DNA from hair follicles (Muir et al. 1994).

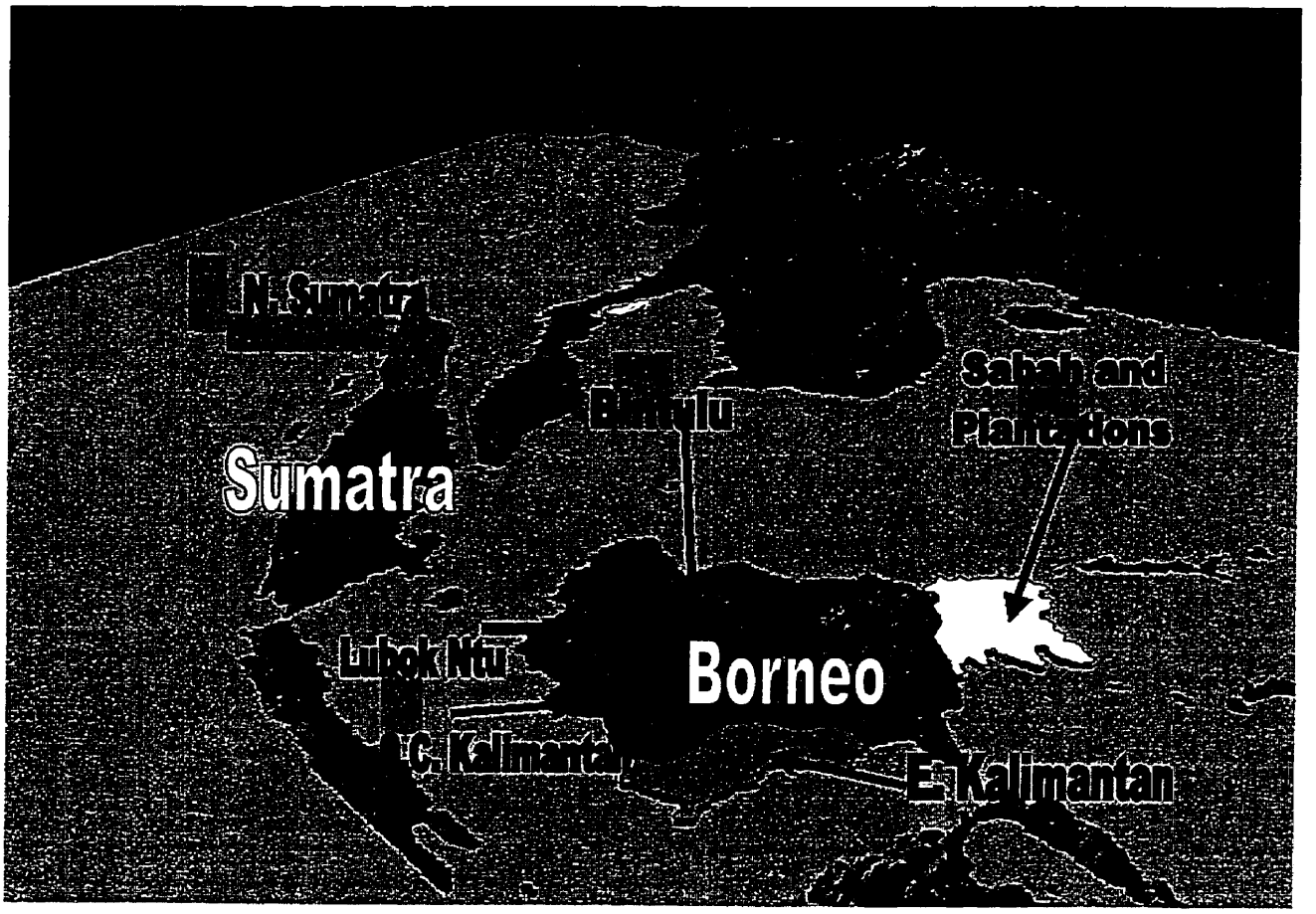


Figure 3: Sample sites are identified

General PCR and Sequencing Methods

Polymerase Chain Reaction (PCR) Whole genomic extracts were used for PCR (Mullis and Fuloona 1987) reactions. In general, less than 50 ng of DNA was used per reaction but individual extracts were not quantified due to the limited availability of

samples. PCR reaction mixtures included 1 mM $MgCl_2$, *Taq* buffer (50 mM KCl, 10 mM TRIS-HCl, .1% Triton X-100), 0.2 mM of primer (Appendix 3), 1 U *taq* polymerase (Bio/Can) PCR reactions followed 35 repetitions of melting at 95 ^{0} C for one minute, annealing at 60 ^{0} C for one minute, and extending at 72^0 C for one minute. Following the 35 repetitions a five minute extension was used.

Sequencing PCR amplification primers were used for sequencing. ^{33}P labeled dideoxy terminator sequencing (Amersham) was used for direct manual sequencing from Prep-a-gene (Bio Rad) purified PCR products following manufacturers' suggested protocols (cycle sequencing reaction over 40 cycles of 90° C/30s, 55° C/30s, 72° C/30s). Sequence reaction products were separated on 6% urea polyacrylamide gels in glycerol tolerant (taurine) buffer. Sequence alignment was done manually with ESEE3s software (Cabot and Beckenbach 1989). A list of primers used for PCR and sequencing is given in Appendix 3.

Chapter 4

tRNA's

Along with other loci characterized for this study, the DNA sequence was determined for four mitochondrially encoded tRNA's. This chapter includes aligned sequence for the mitochondrially encoded tRNA genes for arginine, glycine, proline, and lysine of the individuals listed in Table 2. These data were not used for phylogenetic analysis because there are insufficient number of informative sites. Nevertheless, the haplotypes determined in these tRNA sequences are consistent with the phylogenetic results from other loci analyzed for this study.

The tRNA genes showed a range of variation. For example, there is a single haplotype for the proline gene shared by all individuals sequenced. However, the genes for glycine and arginine both have haplotypes which are shared by a range of individuals from 50% of the samples to 2.5% (one individual) of the samples. Some of the well represented haplotypes found have sequences that appear to have diminished secondary structural stability. These unique genotypes are a likely result of population sampling. Even if selection is an important influence on the stability of a given genotype in a population, the effect of selection may not have been sufficient to preclude drift in these small populations. In fact, there does not appear to be a correlation between haplotype frequency and inferred secondary structure stability.

Methods

Sample sources, DNA extraction, PCR and sequencing methodologies are outlined in Chapter 3. Primers used for amplification and sequencing are found in Appendix 3. The individuals sequenced for each tRNA gene is summarized in Table 1.

Table 1: Names and origins of all individuals sequenced for tRNA's.

Name	Origin		Arg	Gly	Pro	Lys
Bella	Sumatra		X			X
X97707*	Sumatra		X	X	X	X
Ruby	Sumatra		X	X		
Abigail	Sumatra					X
D38115*	Borneo		X	X	X	X
Rosemary	S.Central Kal.		X	X		
Maggie	Bintulu		X			
Hobler Lily	S.Central Kal		X			
Davida	Borneo		X			
Hing Lee	Sabah		X	X		
Semenduh	Sabah		X			
H2	Sabah		X			
V03	Sabah		X			
V04	Sabah		X			
X01	Sabah		X			
W03	Sabah		X	X		
W05	Sabah		X			
W02	Sabah		X			
Tim	Sabah		X	X		
Supinah	W. Kalimantan		X			
GensusuliSabah		X				
Kim Long	Sabah		X			
Apollo Bob	W. Kalimantan		X	X		
CLO248	Borneo		X	X		
Mellie	E. Kalimantan		X	X		
Baboon	W. Kalimantan		X			
Julie	S.Central Kal		X			
Diane	Borneo		X			
Kelly	Borneo		X			
Patti	W. Kalimantan		X			
Roger	W. Kalimantan		X			
Siswi	W. Kalimantan		X			
Mark	S.Central Kal		X			
Brook	S. Central Kal		X			
Herbie	W. Kalimantan		X	X		
Stan	S. Central Kal		X			
Doc	Borneo		X			
Bebeta	W. Kalimantan		X			
Lemot	Lubok Ntu		X			
CLO1239Borneo		X				
WO7	Sabah		X			
AH FONG	Sabah			X		
ROBOT	TP			X		
RANTO				X		
TEROGON			X	X	X	
RIA					X	
GENTUTU					X	
KIM LONG					X	

*** = GenBank sequences (D38115 from Horai et al. 1992, X97707 from Xu and Arnason 1996)**

Results

Arginine

The gene for the arginine tRNA is found 3' of the gene for ND3. Sequence was determined for 41 individuals (Table 1). There are four variable sites in the gene for $tRNA^{arg}$ (Fig 3 a). These changes include three that do not affect structural integrity and one interruption of base pairing in the TΨC stem (Table 2). Maximum pairwise difference between any individual is a single change (1.5%) over 66 bp. The predicted secondary structure for the tRNA is given Fig. 4 a.

Table 2: The position and nature of all changes for $tRNA^{Arg}$.

position #	changestructure location	putative base pairing position?	change in base pairing?
7	A-G AA stem	yes	no
42	A-G var loop	no	no
47	U-G TΨC stem	yes	yes*
52	G-A TΨC loop	no	no
52	G-C TΨC loop	no	no

* GA base pairing has been shown by Steinberg and Cedergren (1995)

Glycine

The gene for the glycine tRNA is found directly 5' of the ND3 gene. The 15 individuals characterized for the glycine tRNA fall into two haplotypes. The GenBank sequence D38115 plus 13 other individuals share a single haplotype and X97707 (GenBank) has a unique haplotype that differs at four sites over 68 bases (Fig. 3). Three of the changes occur in non-base pairing positions while a C-U transition in the TΨC stem results in an interruption of base pairing (Table 3). The predicted secondary structure is given in Fig. 4 b.

Table 3: The position and nature of all changes for $tRNA^{gly}$.

position #	changestructure	putative location	putative base pairing position?	change in base pairing?
16	G-A	DHU loop	no	no
41	C-U	var loops	no	no
54	G-A	TΨC loop	no	no
56	C-U	TΨC stem	yes	yes

Proline

The gene for the proline tRNA is located directly 5' of the control region. There were no differences found among six individuals sequenced for the proline tRNA locus including X97707 (Fig. 3 c). Predicted secondary structure for the proline tRNA is given in Fig. 4 c.

Lysine

The gene for lysine tRNA is found 5' of the COII gene. There is a 41 base pair spacer found between the genes for $tRNA^{lys}$ and COII. The DNA sequence for $tRNA^{lys}$ was determined for seven individuals. The sequences were found to be variable at five positions over 70 in total (Fig. 3 d). Mutations were found in the AA stem, DHU loop, and the TYC loop, and stem (Table 5). Three of the mutations occur in base-pairing positions (Table 4). The predicted secondary structure is shown in Fig. 4.

Table 4: The position and nature of all changes found for $tRNA^{lys}$.

position #	change	putative structure location	putative base pairing position?	change in base pairing?
2	U-G	AA Stem	yes	no?
14	U-A	DHU loop	no	N/A
44	U-A	TΨC stem	yes	yes
50	T-A	TΨC loop	no	N/A
66	T-C	AA stem	yes	yes

a) tRNA^{arg}

```
              AA           DHU             AC                   TΨC
             stem         loop            loop                  loop
            |>      <|  |><|>  <|> <|  |>  <|>|||  <|> <|     |><|>  <|><|>
                                                                         60
MELLIE      UGGUAAAUAGUUUAAACAAAACAAAUGAUUUCGACUCAUUAAAUUAUGACAGCCAUAUUU
BELLA       ...................................................A........
GENSUSULI   ..............................................G.............
MAGGIE-1    ......G.....................................................
HINGLEE-1   ...................................................C........
JULIE       .........................................G..................
             AA
            stem
              <|
               66
MELLIE      ACCAAA
BELLA       ......
GENSUSULI   ......
MAGGIE-1    ......
HINGLEE-1   ......
JULIE       ......
```

b) tRNA^{gly}

```
               AA          DHU                AC                    TΨC
              stem         loop               loop                  loop
            |>     <|  |><|>  <|> <||>   <|> |||<|>    <|    |>  <|>  <|>  <|
                                                                           60
APOLLO BOB  ACUCUUUUAGUAUAAGCAGUACCGUUAACUUCCAAUUAACCAGUUUUGACAACACUCAAA
X97707      ...............A........................U............G.C....
                  68
              AA
             stem
            |>   <|
APOLLO BOB  AAAGAGUA
X97707      ........
```

c) tRNA^{pro}

```
    AA            TΨC                 AC                DHU           AA
   stem           loop                loop              loop          stem
                                                                           70
  |>    <|   |> <|>   <|>  <|>    <|> |||<|>    <|> <||>    <|> <|   |>    <|
UAGUCUCUUUUUCGUGAAUUGAAGUGGUAGUCGGGGGUUUCGGUUGUAAGAUUAAAAUUUGAUGAGAGAC
```

d) tRNA^{lys}

```
             AA    DHU  DHU    DHU    AC     AC     AC    TΨC   TΨC    TΨC    AA
            STEM   STEM LOOP   STEM   STEM   LOOP   STEM  STEM  LOOP   STEM   STEM
                                                                               70
            |     |    |    |     |     |  ---  |     |     |      |      |     |
D38115      GUAACGUUUCGAUUGGAUCGUAAUUGGAAAAUUCAAUUUCUGAUUCTCUUGGUCGGAGAGAAACGUUACU
Bella       ...........................................A..........................
Abigail     .G....................................................................
X97707      .G...........A...................................A...............C....
```

Figure 4: a) Nucleotide sequence for tRNA^{arg}. In total, 41 individuals were sequenced for this locus. Mellie represents 21 individuals (including X97707 from Sumatra), Gensusuli represents 14, Bella represents two, and Hing Lee represents two. Julie and Maggie are unique sequences. b) Nucleotide sequence for tRNA^{gly} for 15 individuals. The sequences are of two types, Apollo Bob, which represents 14 individuals (including D38115), and X97707 that is unique. c) Nucleotide sequence for tRNA^{pro}. All sequences, including X97707 and D38115, were identical. d) Sequence alignment for tRNA^{lys}.

Chapter 5
ND3

Introduction

Until about thirty thousand years ago the orangutan (*Pongo pygmaeus*) was found throughout South Asia (Smith and Pilbeam 1980, Temerin 1980, Andrews and Cronin 1982) and presumably had significantly larger population size than today. Although contemporary populations are arboreal, some have suggested that the orangutan also inhabited a terrestrial habitat prior to the Wisconsin glaciation (Smith and Pilbeam 1980). The evidence has been interpreted as equivocal on this point (Temerin 1980, Galdikas 1981). The extirpation of the orangutan from South Asia was probably influenced by paleoclimatological and paleoecological changes brought on during a period (~35 kya to ~9 kya) of the Wisconsin glacial epoch. Dramatic shifts in ecological structure (Vander Kaars 1990, Adams 1992) preceded the southward advance of the glacial front. Sequestering of water, in combination with sheer mass of the advancing glacier had important effects on coastal sea levels and land bridge formation (Muir et al. in review a & b). At the Last Glacial Maximum (LGM) (~15 kya) sea level was ~125 meters lower than contemporary sea level.

Paleoecological predictions based on percussion core data correlate paleoecological change with the advancing Wisconsin glacier (Vander Kaars 1990, Vander Kaars and Dam 1996). Orangutan populations on the South Asian mainland probably expanded south following an abundant food supply. Persistence of a connecting land bridge resulted in its transformation to tropical woodland or jungle (Adams 1993 or see http://www.esd.ornl.gov/ern/qen/refs.html). This open niche may have offered tempting forage for the orangutan. If this land bridge had recently provided a habitat for

the orangutan we might expect to see evidence of a particular distribution of haplotypes.

Sequence data for the mitochondrial-encoded gene, ND3, is presented for 37 Bornean and four Sumatran orangutans. The analysis of these data was undertaken to interpret contemporary population structure. Simulations of previously available habitat were used to test a proposed population structure based on analysis of ND3 sequence data. The amount of sequence variability is high especially among individuals from Sumatra. It appears that the orangutan populations in Sumatra and Borneo have significant sub-structure as sequence divergence of the level seen in Sumatra is indicative of long periods of genetic isolation.

The long-term survival of the orangutan may depend on understanding population structure, historical patterns of dispersal, and the maintenance of diversity.

Methods

All individuals sampled for this study are identified in Table 5 with their geographic origin listed to the precision available. DNA preparation, PCR and sequencing are described in Chapter 3. Haplotypes that are encountered more than once are designated by the name, all lowercase, of one of the individuals which carries the haplotype (Table 5). Individuals identified alpha-numerically are of three groups. CLO1248, and CLO1239 were collected from rehabilitant orangutans from near to Tanjung Puting, while D38115, and X97707 are identifiers of GenBank sequences. The remainder of alphanumerically named orangutans were individuals who were translocated from palm oil plantations to the Tabin wildlife reserve.

Table 5: Names, origins, and haplotype designations for all individuals sampled for this study.

Name	Origin	Haplotype Designation
Bella	Sumatra	Bella
X97707*	Sumatra	X97707
Ruby	Sumatra	maggie
Abigail	Sumatra	maggie
D38115*	Borneo	maggie
Rosemary	S.Central Kal.	maggie
Maggie	Bintulu	maggie
Hobler Lily	S.Central Kal	maggie
Davida	Borneo	hinglee
Hing Lee	Sabah	hinglee
Semenduh	Sabah	tim
H2	Sabah	tim
V03	Sabah	tim
V04	Sabah	tim
X01	Sabah	tim
W03	Sabah	tim
W05	Sabah	tim
W02	Sabah	tim
Tim	Sabah	tim
Supinah	W. Kalimantan	tim
Gensusuli	Sabah	tim
Kim Long	Sabah	tim
Apollo Bob	E. Kalimantan	apollobob
CLO248	Borneo	apollobob
Mellie	E. Kalimantan	Mellie
Baboon	W. Kalimantan	Baboo
Julie	S.Central Kal	Jul
Diane	Borneo	Diane
Kelly	Borneo	Kelly
Patti	W. Kalimantan	Patti
Roger	W. Kalimantan	Rogr
Siswi	W. Kalimantan	Siswi
Mark	S.Central Kal	Mark
Brook	S. Central Kal	Brook
Herbie	W. Kalimantan	Herb
Stan	S. Central Kal	Stan
Doc	Borneo	Baboo
Bebeta	W. Kalimantan	Bebta
Lemot	Lubok Ntu	Lemot
CLO1239	Borneo	239
WO7	Sabah	W07

* = **GenBank sequences** (D38115 from Horai et al. 1992, X97707 from Xu and Arnason 1996)

Results

The mitochondrial gene ND3, which comprises 345 bases, was sequenced for 38 individuals and two sequences obtained from GenBank were added to the comparison (Appendix 4). Accession D38115 is from Borneo and shares a haplotype with maggie, and accession X97707 is from Sumatra.

The greatest sequence divergences between individuals are found within Sumatra and between Sumatra and Borneo. Nevertheless, variation between individuals within Borneo is also substantial (see Table 6). There are 20 distinct haplotypes in the 37 individuals sampled from Borneo (Fig. 5).

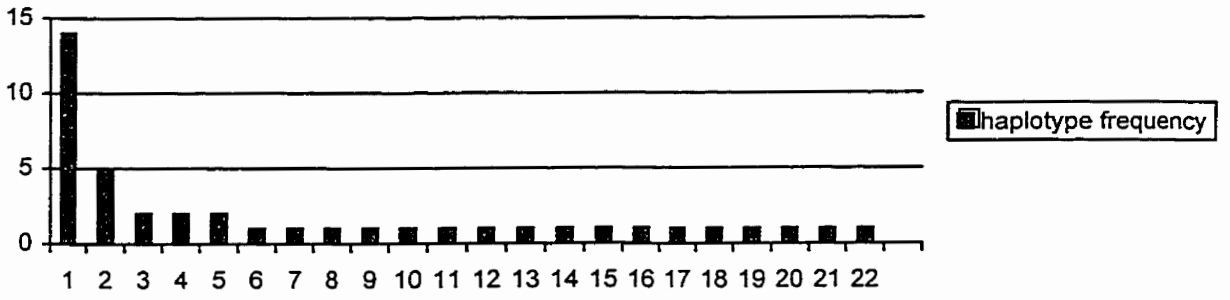


Figure 5: Haplotype frequencies are given for each of 22 haplotypes found in Borneo and Sumatra. Haplotypes 1-20 were found in individuals from Borneo. Haplotypes 2, 21, and 22 were found in the three Sumatran sequences.

Of these 20 haplotypes, five are encountered more than once. The most frequently encountered haplotype, tim, has 14 constituent individuals (Fig. 6 haplotype α). There are two individuals (Davida and HingLee) which share the hinglee haplotype. Apollo Bob and CLO1248 share the haplotype apollobob, and Doc and Baboon share the "baboo" haplotype. The only other haplotype found in more than a single individual is

maggie (D38115, Ruby, Rosemary, Abigail, Maggie, and HoblerLily: Fig. 6 haplotype θ). The maggie haplotype has the most widely distributed range and is particularly interesting as it was found in individuals from both Borneo and Sumatra. The sharing of a haplotype between individuals from Borneo and Sumatra would not be expected under the argument given by Xu and Arnason (1996) and Zhi *et al.* (1996) and argues strongly against a suggested Sumatra/Borneo species split (Xu and Arnason 1996; but see Muir *et al.* 1998).

Pairwise comparisons of individuals from Borneo show up to 10 differences (2.9%) most of which are silent but include two replacement substitutions. Within Sumatra comparisons show from 17 to 26 (4.9% to7.5%) differences including up to 21 silent changes. The pairwise comparisons between individuals from Borneo (37 sequences) and those from Sumatra (4 sequences) however, show divergences up to 8.1% (or 28 differences) (Table 6). Note that the observed 0% divergence is not within Sumatra but between two Sumatran and four Bornean orangutans. Of the 13 replacement substitutions found in this data set, five are conservative changes while the remainder are non-conservative.

The sequences for three individuals from Sumatra were determined (Abigail, Bella, and Ruby) and compared to X97707. The three Sumatran haplotypes fall into three distinct groups, which are all quite diverged from each other.

Figure 6: The distribution of haplotypes is shown on a map of Borneo and Sumatra. Pies show the frequency of each haplotype within its population. Haplotypes that were found in more than one population are identified in Greek letters. The geographic origins for eight individuals used in this study cannot be identified more specifically than Borneo. Records for the origin of Sumatran samples are similarly vague. Map is adapted from Muir et al., a in review.

Discussion

Sampling

A major difficulty encountered with studying molecular variation in orangutan populations is the acquisition of samples and the verification of their geographic origins. Although captive orangutans are plentiful (913 living orangutans are currently registered in the International Studbook of the orangutan 1995), most of these individuals are second or third generation captive. Records as to the exact location of

Table 6: Comparisons are given for replacement substitutions (# repl. sub.), silent substitutions (# sil. sub.), silent to replacement ratio (sil./rep.), total number of variable sites (# var. sites), and transition to transversion ratio (s/v). Comparisons are made among orangutans from within Borneo, within Sumatra, between Borneo and Sumatra, between the reference sequence and *Pan paniscus* (Ppa), and between the reference and *Gorilla gorilla* (Ggo)

	# repl. sub.	# sil. sub.	si/repl	# var. sites	% variation	s/v
Borneo	0 - 2	0 - 8	0 - 4	0 - 10	0 - 2.8	0 - 2
Sumatra	3 - 5	15 - 21	4 - 5	17 - 26	4.9 - 7.5	4 - 17
Bor/Sum	0 - 4	0 - 25	0 - 6.3	0 - 29	0 - 8.1	0 - 4.8
maggie/Ppa	25	35	1.4	61	17.6	3.4
maggie/Ggo	23	33	1.4	56	16.2	3.3

their ancestors' captures are sparse if available, and not available in many cases. The current system of record keeping, the SPARKS system, by contrast, is detailed but does not add to information available for individual origins (L. Perkins pers. comm.). A complete discussion of population genetics and structure requires information about the origin of the samples used and so samples from wild orangutans are required. This is problematic in the case of the orangutan. The orangutan is semi-social and is found, well dispersed, in dense tropical jungle. Furthermore the orangutan is an arboreal canopy dweller and is adept at concealing themselves (Galdikas 1985 a and b). While obtaining hair samples from ground nests has been undertaken for the gorilla (Garner and Ryder 1995), and chimpanzee (Morin et al.. 1993, and Morin et al.. 1994), orangutan nests, which tend to be found in the canopy, are virtually inaccessible without proper rigging.

Samples used in this study are of three types: blood, hair, and feces. All blood samples used for this study were taken during routine physical examinations of captive individuals or wild-born individuals being rehabilitated to the wild. Good records as to

the origins of these individuals are available for 11/18 of these samples. Hair samples were collected from individuals being rehabilitated at the Sepilok and Semingoh Rehabilitation Centres. A few of these records are specific only to regions within Borneo while the balance is detailed. Hair samples were also collected from a number of wild caught translocated orangutans. Detailed records are available for all of these individuals. Fecal samples were collected from only one individual (Tim) used in this study. Tim was a wild caught, and translocated, orangutan for which detailed records are available. Hair samples were also taken from Tim.

Nature of the Variation

There is a high level of divergence found between haplotypes in this study. Variation between individuals in the same sampled area ranges up to 2.8%. The level of diversity of ND3 within Borneo is approximately equal to that found for the mitochondrial control region for the chimpanzee *Pan troglodytes schIinfurthii* (Goldberg and Ruvolo 1997, Ruvolo et al.. 1994). The control region is usually assumed to have a much higher rate of evolution than protein coding genes. These observations are consistent with the suggestion that there is an elevated rate of evolution in the mtDNA of orangutans with respect to other primates (Muir et al. 1994 and 1995). The variation encountered within the Sumatran samples (5.2% - 7.5%) is particularly interesting and is suggestive of extended periods of isolation. The variation found within Sumatra is almost as high as between Sumatra and Borneo, which is about two and a half-fold higher than within Borneo.

Within Borneo, the most divergent pairs are Roger and Herbie, and Roger and Stan. Both pairs show ten differences between them with four to five transversions.

Individuals from Sumatra are all highly divergent from each other. Ruby is 17 sites (with two transversions) different from Bella and differs at 26 sites (with five transversions) from the GenBank sequence X97077. Bella differs at 27 sites (including three transversions) from the GenBank sequence X97077.

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; 2) a cladogram or 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 Tempelton 1993).

In a pairwise comparison of differences sufficient sequence intermediates exist so that a cladogram network can be constructed with internal nodes occupied by contemporaries of the branch tips (Fig. 7). The cladogram of relationships between haplotypes is constructed by first joining haplotypes that differ by a single nucleotide. Haplotypes which are different by 2, and 3 nucleotides are then joined. The cladogram was also checked to confirm that pairwise differences between haplotypes agree with branch length for both adjacent and distant nodes.

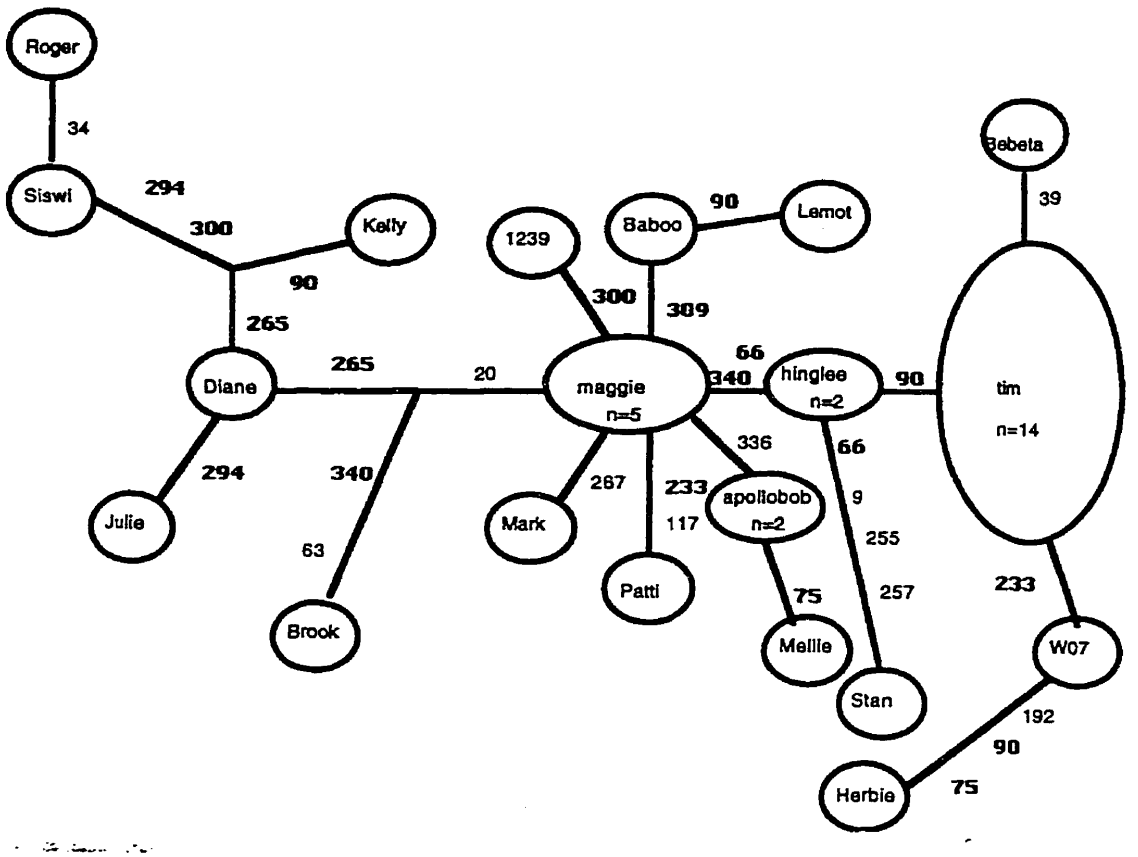


Figure 7: Cladogram of Bornean ND3 haplotypes is shown with the position of branch determining mutations identified above the branch. Numbers printed in bold are sites for which the mutations are inferred to have been multiply and independently derived.

Given the high level of variation observed in this study, and the limited number of sites at which mutations would be tolerated in a protein-coding genes, we expect to observe multiple, independent substitutions at some sites. Three different nucleotides were observed at position 336 (Fig. 7), requiring at least two independent substitutions. By identifying the mutations in each transition between nodes, one can distinguish sites at which mutations are not homologous but independently derived. Among Bornean orangutans, the observed variation at 13 of the sites could have arisen by single mutational events. At least two independent mutations are required to account for the

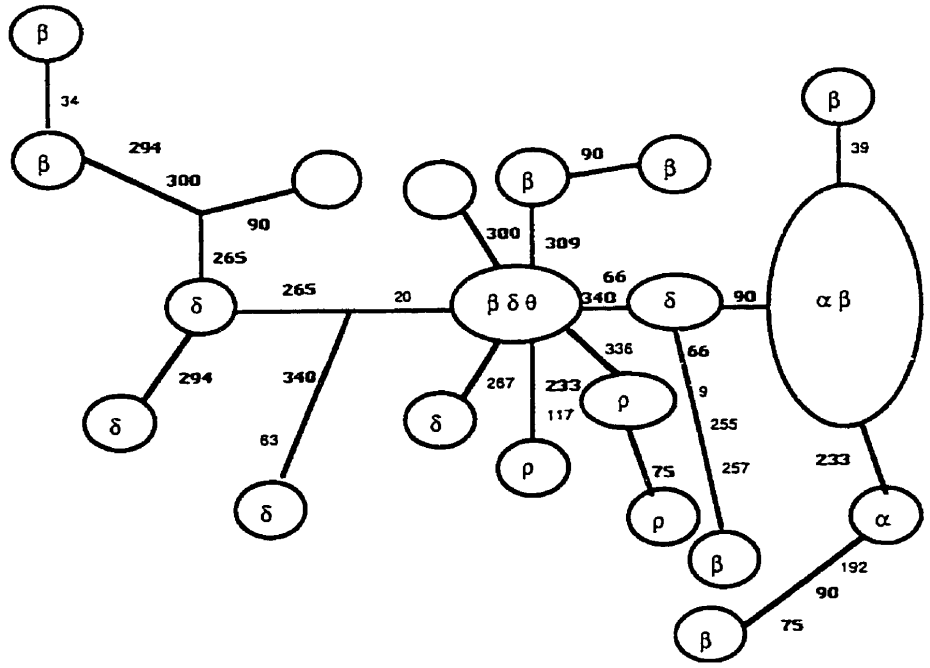
differences observed at each of five sites, while one site (309) requires three, and one (90) requires four independent mutational events. The branch locations of inferred multiple substitutions are shown in bold on Fig. 7. One can, therefore, verify internal consistency of branch lengths by taking into account inferred multiple substitutions.

The network also makes sense geographically from a migrational perspective. All links between haplotypes, for which origins are confirmed, are within the same sample area, or between adjacent sampling areas (Fig. 8). The cladogram network also, therefore, provides an illustration which is more visually intuitive with respect to modeling past dispersal.

Coalescence

In comparing haplotypes it would be useful if one could be distinguished which is ancestral to the largest number of studied haplotypes. Coalescent theory indicates several lines of evidence for doing this. The central haplotype in a cladogram network is the haplotype that is connected to the largest number of other haplotypes (Crandall and Templeton 1993). Maggie is connected directly to seven other haplotypes while hinglee is connected to four and tim is connected to three. Central positioning, in addition to the fact that maggie in the most geographically widespread (discussed above), leads us to infer that maggie is the "ancestral" haplotype. In an effort to remove any bias which might result from sampling which is not completely random, the coalescent theory was

a)



b)

Figure 8: (A) Geographic origins of the individuals in the cladogram are identified according to the map (B).

extended to a codon by codon comparison. I asked the question "Which haplotype is closest to the codon by codon majority rules sequence?" A majority rules sequence was arrived at by comparing each codon for each haplotype to the codon found in all other haplotypes at that position. The codon that is in the majority is taken to be the oldest state. I tabulated (Table 7), for each haplotype, the sum of the expression :

$$\sum_{i=1}^{A} x_i/y$$

A=variable codon number

x_i= # of individuals which share codon x at site i

y= # of haplotypes

The largest value in this tabulation is taken to be the haplotype that is ancestral to the largest number of the haplotypes that I characterized. This distinction goes to maggie agreeing with the conclusions based on geographic distribution and positioning within the network. The next highest value is held by tim which is also widespread and is a haplotypes shared by the highest number of individuals which cannot entirely be explained by sampling.

Table 7: Values for majority rules sequence similarity for all haplotypes. Values are given in the order that haplotypes appear in Appendix 4.

1) 39.711	7) 38.449	13) 38.361	19) 24.521
2) 38.188	8) 38.057	14) 37.405	20) 39.102
3) 39.059	9) 39.493	15) 36.969	21) 18.825
4) 38.928	10) 36.005	16) 38.581	22) 38.536
5) 37.753	11) 37.231	17) 37.230	23) 39.276
6) 38.319	12) 38.971	18) 38.493	

Sumatran individuals characterized for this study fall into three distinct haplotypes which are quite divergent (up to 7.5%) and one of which is shared with a haplotype of Bornean origin (Ruby shares the maggie haplotype). The distribution of

pairwise differences for all samples used in this study is tri-modal (fig 9) which corresponds to comparisons involving the three Sumatran haplotypes.

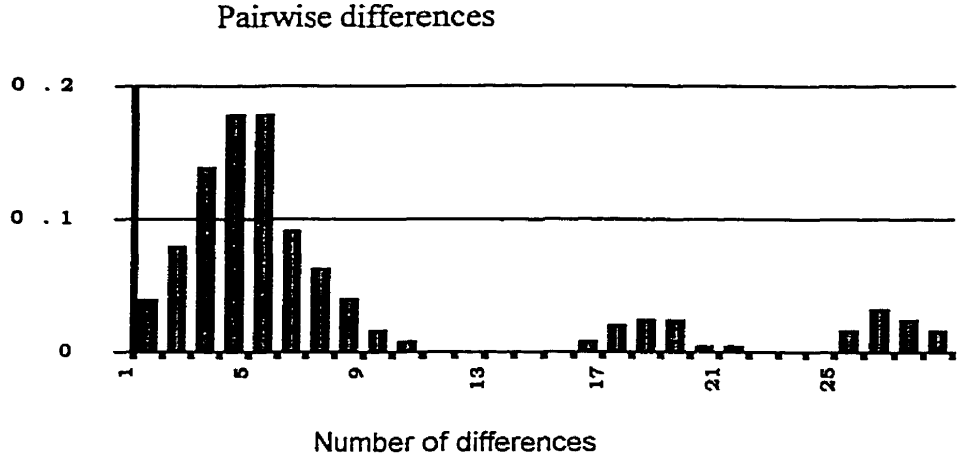


Figure 9: Distribution of pairwise distances, in base pairs, is given for all samples. Middle and right-most nodes are pairwise differences between Bella, and X97707 who are both Sumatran.

Divergence of haplotypes recorded in our study indicates extended periods of genetic isolation. The relationship of the three Sumatran haplotypes is illustrated in Fig. 10. The triangle at the center node of the cladogram corresponds to a mutation at position 39 for which each haplotype has a different nucleotide (Fig.5). Because the sample set for Sumatra is so small I cannot speculate on the overall diversity of haplotypes which may be present in Sumatra although increased size of the sample set could only add to the diversity. The high level of divergence of haplotypes found among Sumatran samples indicates prolonged genetic isolation. A parsimonious explanation of this diversity is that present populations in Sumatra are remnants of at least three previously isolated populations. Persistent isolation of these haplotypes may have been facilitated by a number of east-west running rivers and a north-south mountain range although, since mitochondrial loci do not recombine, contemporary barriers to gene flow

are not required to accommodate recently introduced divergent haplotypes within a population. More samples of known geographic origin are required from Sumatra if we are to fully assess the variation of these populations.

Geographic Distribution

I provide a model to explain present haplotype distribution and the diversity of haplotypes found. The effects of major glacial episodes have resulted in dramatic changes in sea level (-125m at LGM). A drop in sea level of this magnitude results in the emergence of major land bridges between the islands of Borneo, Sumatra, and Java, and the South Asian mainland (Fig 11) (Muir et al. in review a and b). Dispersal routes can be inferred from the presence of this land bridge which are in strong agreement with

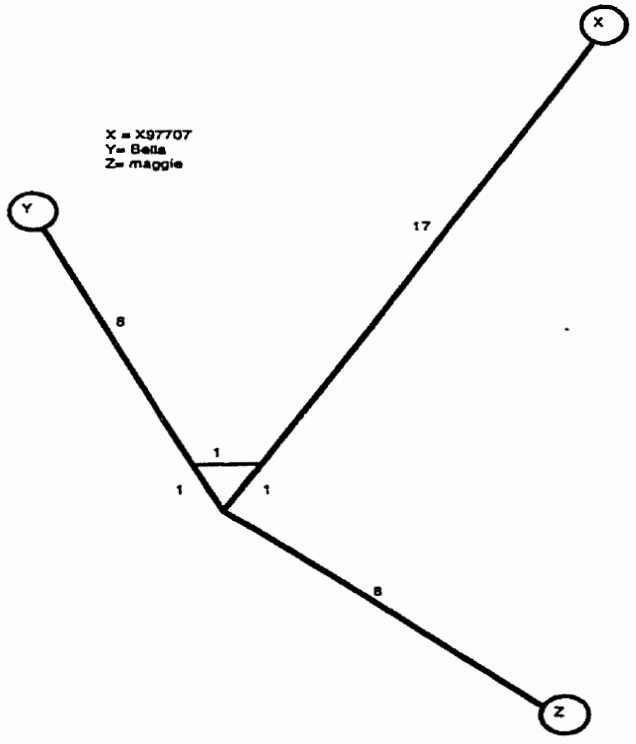


Figure 10: Cladogram for three Sumatran ND3 haplotypes. Triangle at the center is a mutation at position 39 for which the three haplotypes have a different base. X=X97707, Y=Bella, Z=maggie (Ruby and Abigail). Numbers on branches indicate number of pairwise differences.

contemporary haplotype distribution.

The maps shown in Fig. 11 represent a time series over two major glacial epochs. In the first frame, representing a period at least two to three hundred thousand years ago, sea levels are shown to be similar to today. The ancestral distribution of orangutans, through South Asia but prior to colonization of Borneo, Sumatra, and for a period Java, is indicated. As the sea level drops to approximately 40 metres lower than today, frame two shows that land bridges begin to appear between the mainland and the islands. The arrow indicates one of many viable dispersal routes (Drawhorn 1994) and proposed colonization of Sumatra. As the sea level drops to –125 m at glacial maximum, the islands of the archipelago west of Wallace's line become a single landmass joined to the mainland. Piston core analysis indicates this inter-island land mass supported savanah and tropical woodland forests (Van der Kaars 1990, Van der Kaars et al. 1996, Adams 1992 and 1993) which could have supported orangutan populations. The curved arrow in the third frame indicates proposed colonization of the entire land mass by orangutans. As sea levels rise 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 subsequent glacial period(s) opportunity existed for populations, which would have become genetically distinct, to disperse from Borneo and the mainland into Sumatra along routes proposed by Drawhorn (1994) as indicated in frames 6, 7, and 8. As indicated in frame 9 the mainland population of orangutans has been extirpated while Sumatra supports genotypes which are highly diverged, one of which shares

identity with Bornean orangutans. The genotypes on Borneo are much less diverse.

The most widespread haplotype (maggie) is found at sites that border the position, geographically, at which the bridge is most persistent (Fig. 6: haplotype θ and Fig. 11). Since Abigail and Ruby share a haplotype (maggie) with orangutans from Borneo, their ancestors were likely Bornean. The fact that the maggie haplotype is found at sites close to a land bridge, which is currently under only 30 m of ocean, also supports a dispersal hypothesis. I suggest that the X97707 haplotype is a descendant of migrants from presently extirpated populations in mainland south Asia (Fig. 11, and Video 3).

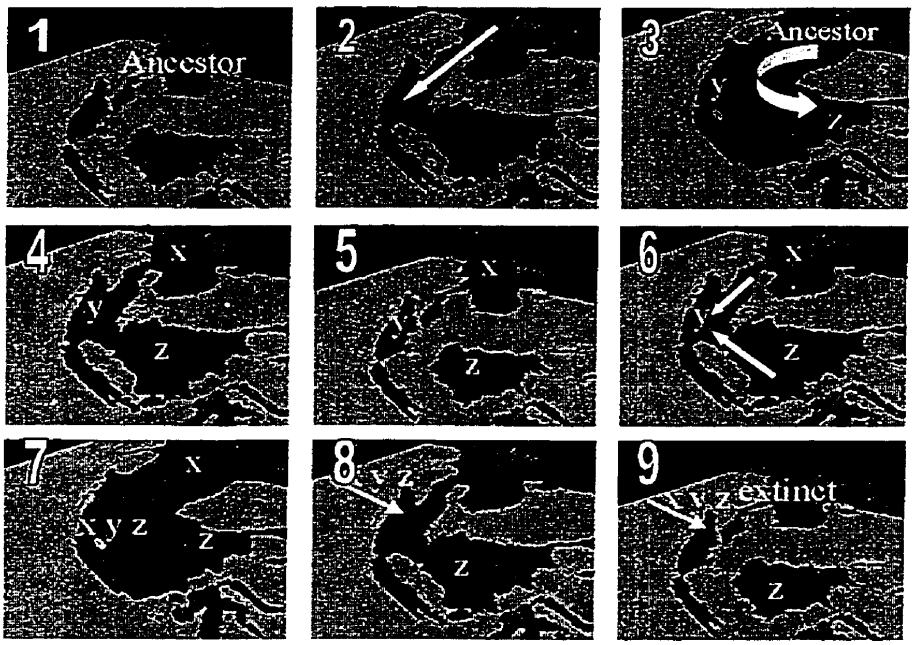


Figure 11: Paleodispersal predicted based on haplotype distribution and availability of route. Maps show the effect of falling sea level on the presence of land bridges. Arrows indicate dispersal possibilities. Map series adapted from Muir et al., in review (a and b) and represents two full glacial advance and retreat cycles and run left to right, top to bottom.

Conclusion

Contemporary distribution of the orangutan is the consequence of migration at

least on some level. Fossil remains suggest a historic distribution throughout South Asia. Data presented here suggests a population structure reflective of a geological history that has been heavily influenced by a number of glacial periods. Since the water is quite shallow in this region, small changes in sea level have had a dramatic effect on the extent of land above water (Fig 11, and Videos 1, and 2). The change in landform has, in turn, provided a number of opportunities for dispersal from the mainland to both Sumatra and Borneo, and between the islands. Genetic diversity of Sumatran individuals indicates extended periods of genetic isolation. Since the pattern of glaciation over the past million years has recurred every 100-150 ky (Adams 1993), repeated opportunity for dispersal has been afforded by the periodic reappearance of the inter-island land bridge.

This hypothesis is further supported by the paleoecological argument given earlier. Based on the data given here I suggest that there has been gene flow between the islands of Borneo and Sumatra. During the last glacial maximum the extirpation of the orangutan from the South Asian mainland is likely to have followed some range expansion south to what is now the island of Sumatra. This dispersal hypothesis is consistent with the observed variation found in Sumatran haplotypes.

There are numerous accounts of the high level of intraspecific variation found in *Pongo pygmaeus* (Ruvolo et al. 1994, Muir et al. 1994 and 1995, Zhi et. al 1996, Xu and Arnason 1996). I have presented data that are consistent, in a variety of ways, with recent migration between currently well defined demes. Putting aside a more philosophical discussion of the phylogenetic species concept (see Muir et al. 1998), a recent migration between populations argues against their delineation as separate species. If there is more than one species of orangutan represented in present populations their

distinction appears not to be a simple bifurcation between the islands of Borneo and Sumatra. Even if the rate of acceptance of mutations in the orangutan is the same as in African Great Apes, it remains probable that dispersal has recently played a role in shaping contemporary population structure. Clearly an increased sample set including samples of known specific origin would benefit understanding of orangutan population structure.

Chapter 6

Cyt b/ψ Cyt b

Introduction

ND3 sequence comparison shows extensive variation and gives substantial evidence of population structure among orangutans (Chapter 5). Within Sumatra, there are at least three highly divergent haplotypes. It is hypothesized that these three haplotypes represent the effects of long-term isolation followed by convergent range expansion or shifts. It is noted that historic, episodic, glacial events resulted in dramatic sea level changes which resulted in the formation of land bridges between the islands of Borneo, Sumatra, Java, and the South Asian mainland (Peltier 1994; Chapter 2). Enormous landmasses that emerged during periods of low sea level opened up vast tropical woodland and Savannah ecosystems (Van der Kaars 1990). It seems likely that the orangutan would have made use of the resources offered by these ecosystems more than just the one time.

It is further noted that Sumatra is geographically situated in a central position between Borneo, and the historic range throughout South Asia (Martin and Andrews 1993). Shared haplotyes found among individuals on Borneo and Sumatra further supports the inter-island migration model as proposed in the previous chapter.

The above conclusions were based on the sequence comparison of a maternally inherited mitochondrial gene. Orangutan social structure obviously plays an important role in allele distribution. Since male and female home ranges may be dramatically different (Galdikas 1985 a and b), there may be sex specific migration between populations or demes. Mitochondrial haplotype may, therefore, be missing half the story.

The purpose of this chapter is to extend the analysis of sequence comparison to a nuclear encoded pseudo-gene. The mitochondrial genome is estimated to evolve at a rate approximately 10 times faster than the nuclear genome for a variety of reasons (Brown et al. 1982). A pseudo-gene was chosen to remove selection as an influence on the rate of evolution and increase the likelihood of the gene being informative at the population level. I chose to analyze the sequences for Pseudo-Cyt b (ψ Cyt b) found on chromosome *II* (Collura and Stewart 1995, see fig 12)). This gene is the result of a mitochondrial duplication which occurred in the ancestor of extant apes approximately 30 million years ago (MYA) and is found in all living apes (Collura and Stewart 1995). The availability of exclusive primers for both the nuclear and mitochondrial copies of the gene make sequence comparison convenient. Inter-specific versus intra-specific divergence is compared. The mitochondrial Cyt b was also analyzed for comparison to the nuclear pseudo-gene and to previous results for mitochondrial ND3. The nuclear gene sequence comparison was used to test the hypothesis that different male and female home range requirements may have influenced allele distribution.

A portion, from position 641 to 1136, of the mitochondrial protein coding gene Cyt b was sequenced for 35 individual orangutans from throughout the Sumatran and Bornean range. ψCyt b was also sequenced for 27 individuals and the evolutionary rates were compared. Phylogenetic analysis of the Cyt b gene is used to further test the possibility of recent gene flow between Borneo and Sumatra.

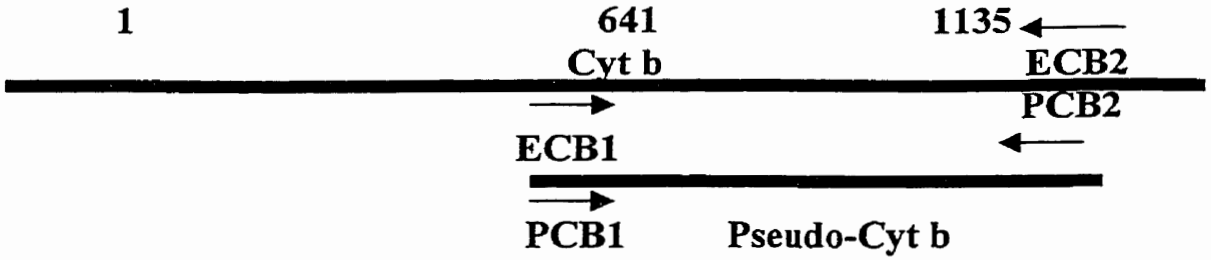


Figure 12: A schematic of Cyt b and the alignment of the nuclear pseudo-gene. Primers used to amplify and sequence these loci are identified.

Methods

Samples sources and general molecular methods are given in Chapter 3. The primers PCB1 and PCB2 (pseudo-Cyt. B), and ECB1 and ECB2 (Cyt b) (Appendix 3) were used for both PCR and sequencing.

Results

The sequence of 495 bases from the 3' end of mitochondrial gene Cyt b was determined for 35 individuals from a variety of locations from across Borneo and Sumatra and were aligned against four sequences previously determined (accession numbers D38115, X97707, X97717, and Dennis; Arnason pers. comm.) Fig. 13). A corresponding segment (419 bp) of a nuclear encoded Cyt b pseudogene was sequenced for 27 individuals and aligned against the GenBank sequence U38269 (Fig. 14). One haplotype of Cytb b was aligned against the GenBank sequence for pseudo-Cyt b (U38269) and was found to have diverged by 138 sites including indels.

Mitochondrial Cyt b

The rate of sequence evolution in Cyt b is in agreement with the rate noted for the mitochondrial gene ND3 (Muir et al. in review c, and Chapter 5). There are ten

haplotypes found among the 33 individuals from Borneo and four haplotypes among the six Sumatran individuals (fig. 14). The maximum sequence divergence found for Cyt b was 8.7% between individuals from Borneo and Sumatra though this divergence was only slightly higher than within Sumatra (8.1%) (Table 10). A striking feature of the pairwise differences between individuals from Borneo and Sumatra is the similarity of some Sumatran haplotypes to their Bornean counterparts. Abigail differs by only three sites or 0.6% from the Bornean reference sequence making it more closely related to Bornean haplotypes than to some other Sumatran ones. Furthermore two Sumatran individuals, Ruby and Gamgar, share sequence identity with 21 Bornean individuals (haplotype SHAGY). Ruby and Abigail also share Bornean sequence identity for ND3 (Chapter 5).

Table 8: A list of individuals who share haplotypes for mitochondrial gene segment of Cyt b.

SHAGY	
1228,	RUBY,
STAN,	HERBIE,
MARK,	SUPINAH,
JULIE,	DIANE,
ROGER,	BROOK,
Y01,	MELLIE,
V04,	GENSUSULI,
KELLY,	HING LEE,
APOLLO BOB,	ANNA,
GAMBAR,	PATTI,
DAVIDA,	SISWI,
HOBLER LILY	X01

WAF
W01,
W05,
AHFONG

```
                                                                      60
WAF         ACTCCGATAA AATCACTTTC CACCCCTACT ATACAATTAA AGACATCCTA GGCCTACTCC
BAJpcb      ?????????? ?????????? ?????????. .C....CC.. ...T..T... ..TT..A.TT
SHAGY       .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... .......... .......... .......... .......... ..........
ANNA        .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
GENEBNK     .T.....C.. ......C... .......... .CT....C.. ...T...... ..........
D38115      .......... .......... .......... .......... .......... ..........
MAGGIE      .......... ....G..... .......... .......... .......... ..........
ABIGAIL     .......... .......... .......... .C........ .......... ..........
KIMLONG     .......... .......... .......... .......... .......... ..........
BEBETA      .T..T..C.. ......C... .....T.... .......C.. ...T...... ..........
Jdura       .......... .......... .......... .......... .......... ..........
DOC         .......... .......... .......... .......... .......... ..........
H5          .?........ .......... .......... .......... .......... ..........
MAWAR       .T.....C.. ......C... .......... .C.....C.. .......... ..........
BELLA       .......C.. ....G.C... .......... .C........ .......... ..........
SUMCYTB     .T.....C.. ......C... .......... .C.....C.. .......... ..........
                                                                     120
WAF         TTTTTCTCCT CGCCCTAATA ACATTAACAC TACTCTCACC AGACCTCCTA AGCGACCCAG
BAJpcb      ..C.C..... .CTTT..... .TTC..GT.. .GT.T..G.. T........G G.T.......
SHAGY       .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... .......... .......... .......... .......... ..........
ANNA        .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
GENEBNK     .......... .......... ...C...... .......... .......... ..........
D38115      .......... .......... .......... .......... .......... ..........
MAGGIE      .......... .......... .......... .......... .......... ..........
ABIGAIL     .......... .......... .......... .......... .......... ..........
KIMLONG     .......... .......... .......... .......... .......... ..........
BEBETA      ..C....... .......... .......... .......... .......... ..........
Jdura       .......... .......... .......... .......... .......... ..........
DOC         .......... .......... .......... .......... .......... ..........
H5          .......... .......... .......... .......... .......... ..........
MAWAR       .......... ........C. ...C...... .......... .......... ..........
BELLA       .......... .......... .......... .......... ...T...... ..........
SUMCYTB     .......... .......... ...C...... .......... .......... ..........
                                                                     180
WAF         ACAACTACAC CTTAGCTAAC CCCCTAAGCA CCCCACCCCA CATTAAGCCC GAATGATACT
BAJpcb      .TC.T..... .....TC... .......AT. ...TG..... ..C...A..A C.G.......
SHAGY       .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... .......... .......... .......... .......... ..........
ANNA        .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
GENEBNK     .......... .......... .......... .......... ......A... ..G.....T.
D38115      .......... .......... .......... .......... .......... ..........
MAGGIE      .......... .......... .......... .......... .......... ..........
ABIGAIL     .......... .......... .......... .......... ......?... ..........
KIMLONG     .......... .......... .......... .......... .......... ..........
BEBETA      .......... .......... .......... .......... .......... ..........
Jdura       .......... .......... .......... .......... .......... ..........
DOC         .......... .......... .......... .......... ........A. ..........
H5          .......... .......... .......... .......... ........A. ..........
MAWAR       .......... .......... .......... .......... ......A... ..G.....T.
BELLA       .......... ......C... .......... .......... .......... ..........
SUMCYTB     .......... .......... .......... .......... ......A... ........T.
```

Figure 13 continued next page

```
                                                                          240
WAF         TCCTATTCGC CTACGCAATC CTACGATCCG TCCCCAACAA ACTAGGAGGT GTAATAGCCC
BAJpcb      .TT....T.. A..T...... T..T.....A .......T.. .T.......C ...C.G....
SHAGY       .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... .......... .......... .......... .......... ..........
ANNA        .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
GENEBNK     .......... .......... .......... .......... .......... ..G.......
D38115      .......... .......... .......... .......... .......... ..........
MAGGIE      .......... .......... .......... .......... .......... ..??......
ABIGAIL     .......... .......... .......... .......... .......... ..........
KIMLONG     .......... .......... .......... .......... .......... ...T...G..
BEBETA      .......... .......... .......... .......... .......... ..........
Jdura       .......... .......... .......... .......... .......... ..........
DOC         .......... .......... .......... .......... .......... ..........
H5          .......... .......... .......... .......... .......... ..........
MAWAR       .......... .......... .......... .......... .......... ..........
BELLA       .......... .......... .......... .......T.. G........C ..........
SUMCYTB     .......... .......... .......... .......... .......... ..........
                                                                          300
WAF         TCATGCTATC TATCCTAATC CTAACAACAA TCCCCGCTCT CCACACATCC AAGCAACAGA
BAJpcb      .TC.AT.... C..T..C..T ...G..GTT. .T..T..A.. .......... ..A.....A.
SHAGY       .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... .......... .......... .......... .......... ..........
ANNA        .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
GENEBNK     .......... C......... .......... ....T..C.. T...T..... ..........
D38115      .......... .......... .......... .......... .......... ..........
MAGGIE      .......... ?......... .......... .......... .......... ..........
ABIGAIL     .......... C......... .......... .......... .......... ..........
KIMLONG     .......... .......... ..T....... .......... .......... ..........
BEBETA      .......... .......... .......... .......... .......... ..........
Jdura       .......... .......... .......... .......... .......... ..........
DOC         .......... .......... .......... .......... .......... ..........
H5          .......... .......... .......... .......... .......... ..........
MAWAR       .......... C......... ......G... ....T..C.. T...T..... ..........
BELLA       .......... C......... .......... .......... .......... ..........
SUMCYTB     ....A..... C......... .......... ....T..C.. T...TG.... ..........
                                                                          360
WAF         GCATAACATT CCGCCCATTA AGCCAATTCC TATATTGACT CTTAATCACC GACCTCCTAG
BAJpcb      ....C.T... T.A....... ..T..G.AT. .G.TC...A. ....G..... .....AT.CA
SHAGY       .......... .......... .......... ....C..... .......... ..........
HOBLERLILY  .......... .......... .......... ....C..... .......... ..........
ANNA        .......... .......... .......... ....C..... .......... ..........
TIM         .......... T......... .......... ....C..... .......... ..........
GENEBNK     .......... T........G .......... .......... T.....G... .....T...A
D38115      .......... .......... .......... ....C..... .......... ..........
MAGGIE      .......... .......... .......... ....C..... .......... ..........
ABIGAIL     .......... .......... .......... ....C..... .......... ..........
KIMLONG     .......... .......... .......... ....C..... .......... ..........
BEBETA      .......... .......... .......... ....C..... .......... ..........
Jdura       .......... .......... .......... ....C..... .......... .......A..
DOC         .......... .......... .......... .......... .......... ..........
H5          .......... .......... .......... ....C..... .......... ..........
MAWAR       .......... T......C.G .......... .......... T.....GT.. .....T...A
BELLA       .......... .......... .......... ....C..... .......... ..........
SUMCYTB     .......... T........G .......... .......... T.....G... .....T...A
```

Figure 13 continued next page

```
                                                                      420
WAF          TTCTCACCTG AATTGGAGGA CAACCAGTAA GCTACCCCTT CATTACTATT GGCCAAGTAG
BAJpcb       CA.....A.. ...CA..... ..G.....TG AAC.G..T.. T.....C... ..A..GA...
SHAGY        .......... .......... .......... .......... .......... ..........
HOBLERLILY   .......... .......... .......... .......... .......... ..........
ANNA         .......... .......... .......... .......... .......... ..........
TIM          .......... .......... .......... .......... .......... ..........
GENEBNK      .......... .......... .......... ....T..... ...C..C... A.........
D38115       .......... .......... .......... .......... T......... ..........
MAGGIE       .......... .......... .......... .......... .......... ..........
ABIGAIL      .......... .......... .......... .......... .......... ..........
KIMLONG      .......... .......... .......... .......... .......... ..........
BEBETA       .......... .......... .......... .......... .......... ..........
Jdura        .......... .......... .......... .......... .......... ..........
DOC          .......... .......... .......... .......... .......... ..........
H5           .......... .......... .......... .......... .......... ..........
MAWAR        .......... ...C...... .......... ....T..... ...C..C... A.........
BELLA        .......... .......... .......... .......... .......... ..........
SUMCYTB      .......... .........G .......... .......... ...C..C... A.........
                                                                      480
WAF          CATCCGTACT ATACTTTACC ACTATCCTAC TCCTTATACC AACCTCTTCC CTGATCGAAA
BAJpcb       ....TA?GA. ...T..CT.T .T...T..?? ?????????? ?????????? ??????????
SHAGY        .......... .......... .......... .......... .......... ..........
HOBLERLILY   .......... .......... .......... .......... .......... ..........
ANNA         .......... .......... .......... .......... .......... ..........
TIM          .......... .......... .......... .......... .......... ..........
GENEBNK      .....AC.T. ......C..T ........T. .A........ .G...T.... ..A.......
D38115       .......... .......... .......... .......... .......... ..........
MAGGIE       .......... .......... .......... .......... .......... ..........
ABIGAIL      .......... .......... .......... .......... .......... ..........
KIMLONG      .......... .......... .......... .......... .......... ..........
BEBETA       .......... .......... .......... .......... .......... ..........
Jdura        .......... .......... .......... .......... .......... ..........
DOC          .......... .......... .......... .......... .....A.... ..........
H5           .......... .......... .......... .......... .......... ..........
MAWAR        .....AC.T. ......C..T .T......T. .A........ .G....C... ..........
BELLA        .......... .......... .......... .......... .......... ..........
SUMCYTB      .....AC.T. G.....C..T ........T. .A........ .G........ ..........

                                  500
WAF          ACTACATACT CAAA??????
BAJpcb       ?????????? ????......
SHAGY        .......... ..........
HOBLERLILY   .......... ..........
ANNA         .......... ..........
TIM          .......... ..........
GENEBNK      ..C....... ..........
D38115       .......... ....TGAACC
MAGGIE       .......... ..........
ABIGAIL      .......... ..........
KIMLONG      .......... ..........
BEBETA       .......... ..........
Jdura        .......... ..........
DOC          .......... ..........
H5           .......... ..........
MAWAR        .......... ..........
BELLA        .......... ..........
SUMCYTB      ..C....... ..........
```

Figure 13: Sequence for the 5' end of the mitochondrially encoded gene Cytochrome b (Cytb) beginning at position 641. Constituents of the haplotypes which are shared are given in Table 1. BAJ PCB is one of the alleles for the nuclear pseudo-gene for Cyt b.

Table 9: A list of individuals who share alleles for ψ Cyt b.

SHERB:
Julie, Bella, X01, W03, KimLong, Mark, Mellie, Herbie, Gensusuli, Siswi

BAJ:
W05, Baboon, 1239, Roger, Gambar, V04, Bebeta, Mike, Kelly, Jara Kong, Doc, Robot

```
                                                                     60
BAJ         TACACAACCA AAGATATTCT AGGTTTAATT TTTCTCCTCC TCCTTTTAAT AATTCTAGTA
DIANE       .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
APOLLOBOB   .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... ..C....... .....-.... .......... .......... ..........
SHERB       .......... .......... .......... .......... .......... ..........
PPY         .......... .-.....-.. .....-.... ..-....... .......... ..........
                                                                    120
BAJ         CTGTTTTCGC CTGACCTCCT GGGTGACCCA GATCATTACA CCTTAGTCAA CCCCCTAAAT
DIANE       .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
APOLLOBOB   .......... .......... .......... .-........ .......... ..........
HOBLERLILY  .......... .......... .......... .......... .......... ..........
SHERB       .......... .......... .......... .......... .......... ..........
PPY         .......... .......... .......... .......... .......... ..........
                                                                    180
BAJ         ACCCTGCCCC ACACTAAACC ACAGTGATAC TTTTTATTTG CATATGCAAT CTTATGATCC
DIANE       .......... .......... .....T.... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
APOLLOBOB   .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... .......... .......... .......... .......... ..........
SHERB       .......... .......... .......... .......... .......... ..........
PPY         .......... .......... .......... .......... .......... ..........
                                                                    240
BAJ         ATCCCCAATA AATTAGGAGG CGTACTGGCC CTTCTATTAT CCATTCTCAT TCTAGCAGTT
DIANE       .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
APOLLOBOB   .......... .......... .......... .......... ......A... ..........
HOBLERLILY  .......... .C........ .......... .......... .......... ..........
SHERB       .......... .......... .......... .......... .......... ..........
PPY         .......... .......... .......... .......... .......... ..........
                                                                    300
BAJ         ATTCCTGCAC TCCACACATC CAAACAACAA AGCATCATAT TTCACCCATT AAGTCAGTAT
DIANE       .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
APOLLOBOB   .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... ....T.T... .......... .......... .......... ..........
SHERB       .......... .......... .......... .......... .......... ..........
PPY         .......... .......... .......... .......... .......... ..........
                                                                    360
BAJ         CTGTTCTGAA TCTTAGTCAC CGACCTATTC ACACTCACAT GAATCAGAGG ACAGCCAGTT
DIANE       .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
APOLLOBOB   .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... .......... .......... .......... .......... ..........
SHERB       .......... .......... .......... .......... .......... ..........
PPY         .......... .......... .......... .......... .......... ..........
                                                                   420
BAJ         GAACAGCCTT TTATTACCAT TGGACAGATA GCATCTA?GA TATATTTCTC TATTATTCT
DIANE       .......... .......... .......... .......... .......... .........
TIM         .......... .......... .......... .......... .......... .....C...
APOLLOBOB   .......... .......... ........G. .......... .......... .........
HOBLERLILY  .......... .......... .......... .......... .......... .........
SHERB       .......... .......... .......... .......A.. .......... .........
PPY         .......... .......... .......... .......A.. .......... .........
```

Figure 14: The DNA Sequence for the portion of the nuclear encoded pseudo-gene Pseudo-Cytochrome b (ψ–Cyt b) which corresponds to the 5' end of the Cyt b gene sequenced.

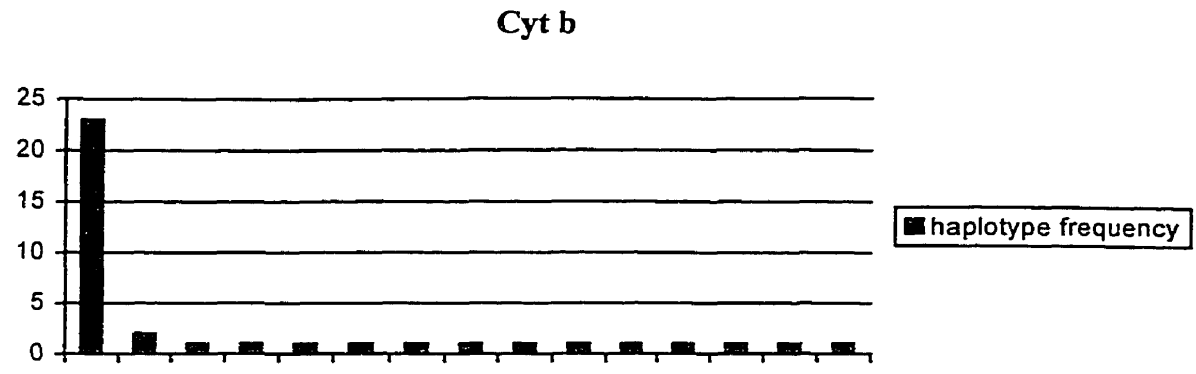


Figure 15: Haplotype frequency for Cyt b haplotypes.

Nuclear ψ–Cyt b

There are eight alleles for the pseudo-gene found among 27 individuals sequenced (Fig 16). The rate of sequence evolution in the nuclear pseudo gene is considerably lower than the mitochondrial protein coding gene. The maximum divergence in ψ-Cyt b is 1.9% between individuals (HoblerLily and ApolloBob) within Borneo, 1.7% between Borneo and Sumatra, and 1.2% within Sumatra. There are an insufficient number of informative sites for phylogenetic analysis.

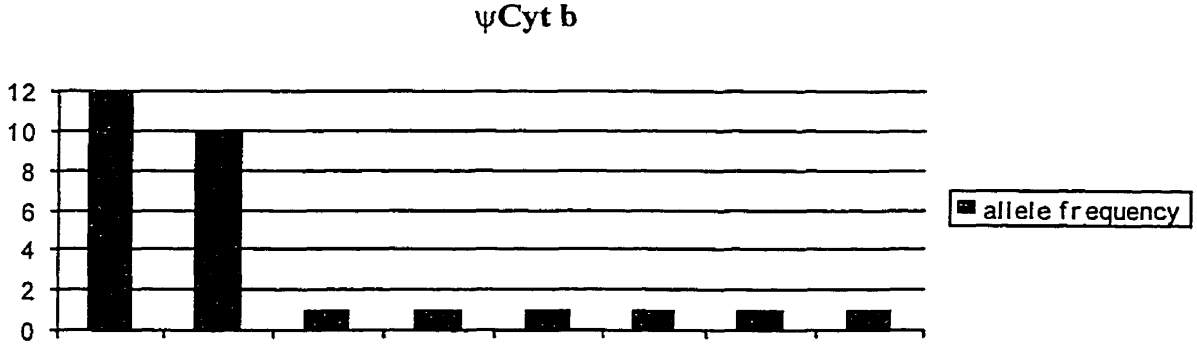


Figure 16: Allele frequency for ψ–Cyt b alleles.

Table 10: Comparisons of divergence, transition/transversion ratios (s/v), and the ratio of silent/replacement substitutions (si/repl)

	Maximum Divergence								
	Within Borneo (Cyt b)	within Sumatra (Cyt b)	between Borneo and Sumatra (Cyt b)	within Borneo (ND3)	within Sumatra (ND3)	between Borneo and Sumatra (ND3)	within Borneo (ψ–Cytb)	within Sumatra (ψ–Cytb)	between Borneo and Sumatra (PCB)
% divergent	2.2	8.5	8.7	2.8	7.5	8.1	1.9	1.2	1.7
s/v	11/3	42/2	43/2	2	17	4.8			
si/repl	12/5	42/14	43/14	4	5	6.3	n/a	n/a	n/a

Discussion

Nature of the Variation

The rate of evolution in Cyt b, as with ND3, among orangutans appears to be elevated with respect to African Great Apes (Garner and Ryder 1994, Morin et al. 1994, Jankowitz et al. 1990, Xu and Arnason 1996, Ruvolo et al.1994, Muir et al. 1994 and 1995, Muir et al. in review c). Although others have explained inter-island diversity by invoking more than a million years of genetic isolation between the islands (Xu and Arnason 1996, Zhi et al. 1996), this hypothesis is inconsistent both with the presence of shared alleles between populations on the two islands, and the enormous sequence diversity within Sumatra.

In addition to a shared rate of evolution, the types of mutations that have been accepted occur at similar rates for Cyt b and ND3. For example, the transition/ transversion (s/v), and the silent/ replacement (si/rep) ratios are similar (Table 11).

Nuclear Vs Mitochondrial

It is convenient, for the purpose of phylogeny construction and analysis of population diversity, that the rate of evolution for all loci is not the same. Different rates at which mutations are accepted result in accumulation of informative sites over different lengths of time. For a given locus to be informative at a particular level of divergence there must be a balance between sufficient, but not excessive, molecular divergence. Comparisons involving individuals that have been reproductively isolated for an extended period of time requires analysis of loci that accumulate mutations more slowly than comparisons of close relatives. Generally speaking nuclear loci evolve more slowly than their mitochondrial counterparts, a characteristic often attributed to improved nuclear error correction (Cann et al. 1984). Interesting examples exist for which geneology of nuclear and mitochondrial loci has been recently shared. There are a number of mitochondrial genes which have been duplicated and transplanted into the nuclear genome (Collura and Stewart 1995, Paabo et al. 1996). If these duplicated "pseudo-genes" (so named because their degenerate sequence is assumed to no longer code for a functional protein) persist in the nuclear genome they will evolve independently from their ancestral gene. Nevertheless the nuclear copy will evolve in a molecular context which is related to the mitochondrial copy (e.g. GC content, secondary structure, slippage potential) The gene/ pseudo-gene example is a good analogy for the phenomenon of bifurcating populations. In both cases, genes are no longer sharing the same pool of alleles and are evolving in different environments.

A problem with the pseudo-gene alignment for distantly related individuals is the need to establish that the pseudo-gene for each individual examined is homologous. As

pointed out by Collura and Stewart (1995) it is not clear that the pseudo-gene for Cyt b that was sequenced for bonobo and chimp is homologous to the one sequenced for orangutan, gorilla, and human. Pairwise comparisons for the great ape alignment put both bonobo and chimp more distantly related to humans than either gorilla or orangutan. According to this alignment, the orangutan is the most closely related ape to humans. Collura and Stewart (1995) note there are two detectable pseudo-genes in the orangutan. Since there need only be a change in the primer annealing region to "lose" a PCR product it is possible that the copies which are aligned do not share a most recent common ancestor.

According to the work of Collura and Stewart (1995) the copy and transfer of a segment of the mitochondrial gene, Cyt b, into the nuclear genome occurred approximately 30 million years ago. The sequence for the nuclear pseudo-gene is aligned for Great Apes, and this alignment is in turn aligned against the population alignment of the sequence for Cyt b for orangutans in Fig. 14.

The rate of acceptance of mutations is considerable higher in mitochondrial than nuclear loci by about seven fold despite the fact that ψ Cyt b is non-coding. It is difficult to assess the differences in the s/v ratio between nuclear and mitochondrial loci since the mutation rates are so different even though the size of the loci are relatively similar. Alleles for both loci have a single (two for ψ Cyt b) predominant type and a number of less well represented alleles (Figs. 14 and 15).

There is a notable difference between the maximum variation found within Borneo and that between Borneo and Sumatra, or within Sumatra for both mitochondrial Cyt b and ND3 (Chapter 5). This observation implies one or more deep splits within the

Sumatran range. The fact that this difference is not apparent in the nuclear locus may, in part, be a result of wider range requirements for males than for females and the possibility of male specific migration between demes which would not be detectable in a maternally inherited locus. Galdikas (1985 a and b) has previously noted that male and female home ranges are significantly different and that males, especially sub-adult males, are sometimes seen to travel in loose sib-packs. Reproductive contribution by these males between demes would result in different allele distribution for nuclear and mitochondrial types if the females remain local.

Phylogenetic relationships among alleles

Gene trees were constructed for Bornean and Sumatran populations using Cyt b data. Pseudo-Cyt b data did not provide sufficient number of informative sites for tree construction. Gene trees were constructed in the form of cladograms that allow internal nodes to be occupied by extant taxa. The cladograms were constructed using the number of differences in a pairwise distance matrix according to Muir et al. (in review c). The Bornean cladogram (Fig. 17) shows a single haplotype, SHAGY, to which most other haplotypes are joined directly by a short branch. The mutations are mapped directly onto the branches between nodes and a mutation which is inferred to have arisen two times independently is shown in outline font. The cladogram for Sumatran types is given in figure 19 that shows three or four distinct haplotypes.

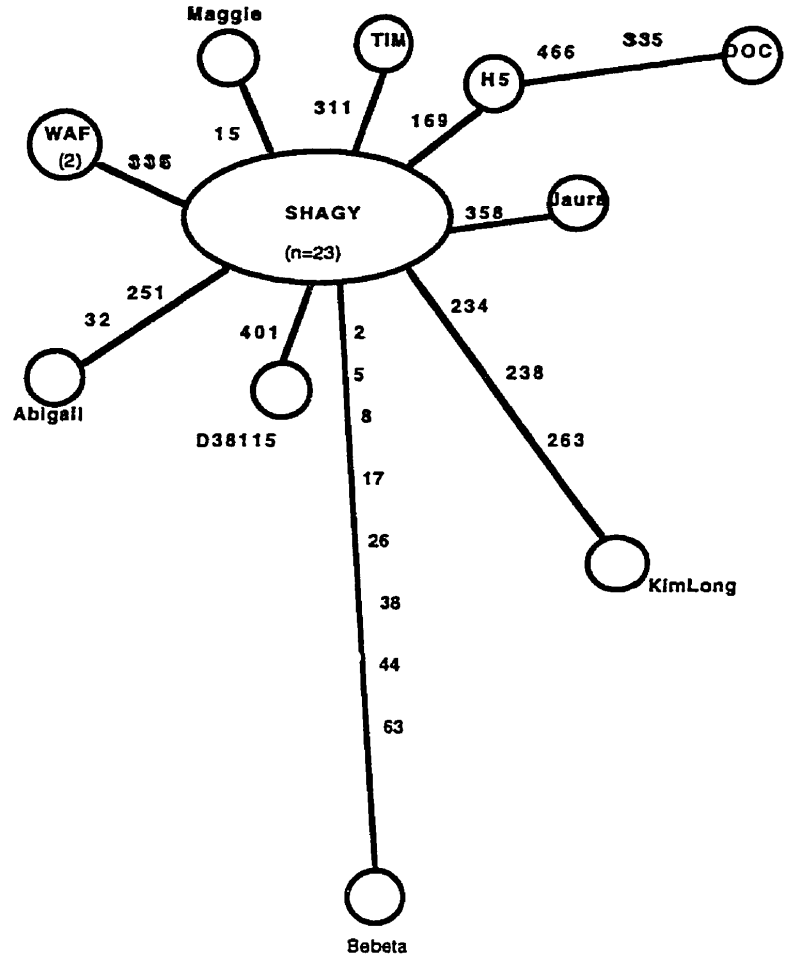


Figure 17: Cladogram of Cyt b haplotypes from Borneo with two (Abigail and Jaura) from Sumatra. In parentheses are the number of constituents which share the SHAGY (23) and WAF (2) haplotypes. Along each bar connecting the haplotypes are the position of each mutation which separates the haplotypes. The mutation at position 335 (in outline) is inferred to have arisen independently twice.

Coalescence

Estimating among aligned sequences which is the most ancestral provides a tool for aiding the construction of a migratory hypothesis to explain allele distributions(Chapter 5). Coalescent theory provides a theoretical framework for

determining which sequence is most like the ancestral state. According to Crandall and Tempelton (1993), the haplotype that is connected to highest number of haplotypes in a cladogram is most ancestral. In the case of the cladogram for Cyt b, the most connected sequence is SHAGY which is also the most geographically wide spread haplotype; another character diagnostic of ancestral affinity. Muir et al. (in review c) also developed a test for ancestral state of protein coding genes (see Chapter 5). This formula is based on the principle that the sequence that is a collection of the highest number of most widely used codons is most like the ancestral state. Following this formula SHAGY is again determined to be most like the ancestral sequence. This determination is consistent with the results of the coalescent diagnosis completed for ND3 (Chapter 5).

Geographic Distribution

As with the case of ND3 (Chapter 5), Cyt b haplotype distribution indicates some demes have recently shared gene flow (fig. 18 and fig. 8). A striking example of this gene flow are haplotypes which share identity or a high degree of similarity and are found separated on the islands of Borneo and Sumatra. Ruby and Gamgar, who are both of Sumatran origin or maternal lineage, share sequence identity with 21 Bornean orangutans.

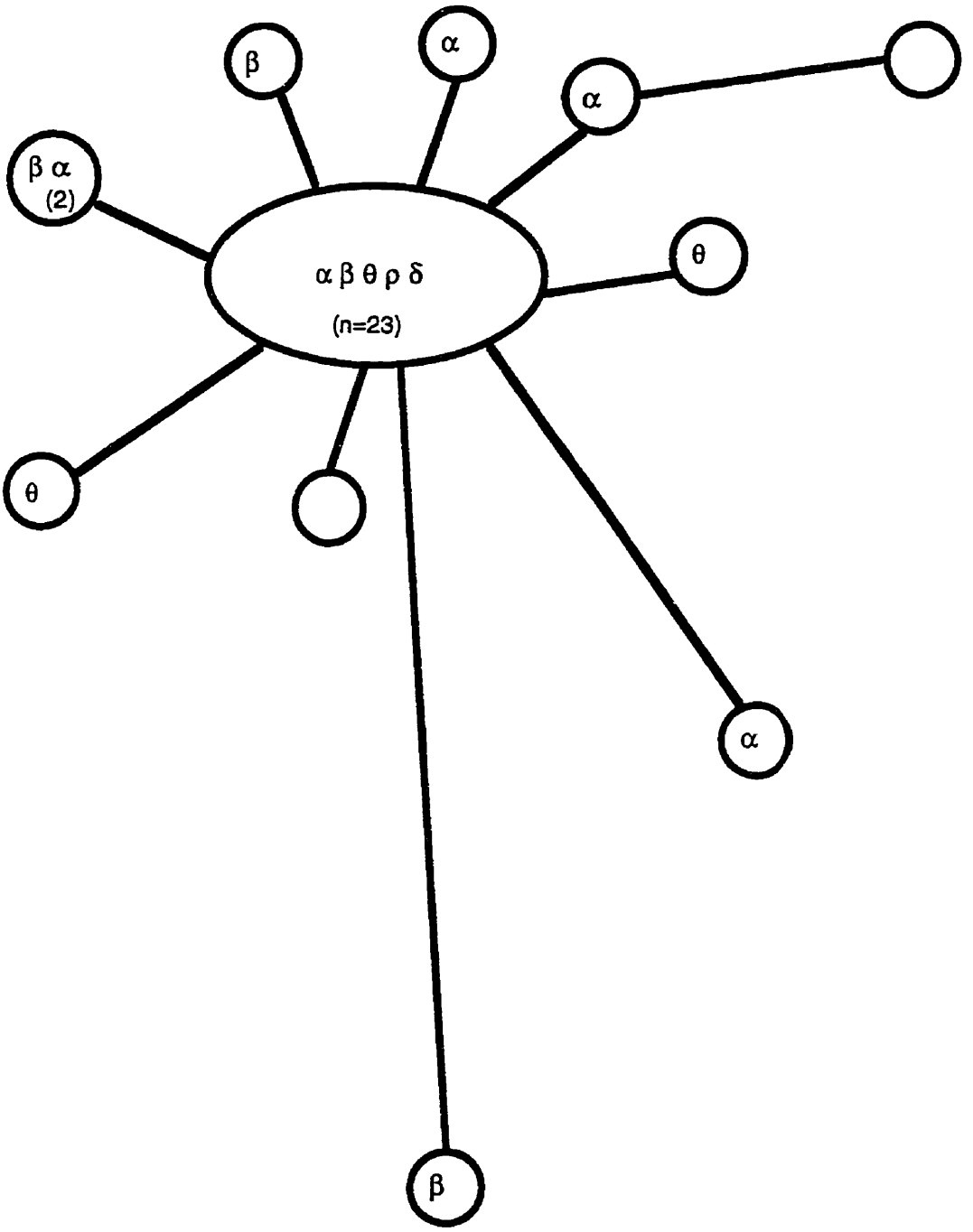


Figure 18: The sample location of Bornean Cyt b haplotypes (Figure 17). Symbols indicate geographic locations identified in Fig. 8 Two haplotypes of individuals for which the origin cannot be verified are left blank.

Moreover, the Sumatran orangutans Abigail, Bella, and Jaura are also more similar to Bornean orangutans than they are to Mias or X97707, which are of Sumatran origin. The presence of sequence identity between some individuals from Borneo with some Sumatran individuals, contradicts the assertion that the populations of Borneo and Sumatra could be considered separate phylogenetic taxa (Ryder and Chemnick1993, Xu and Arnason 1996, Zhi et al. 1996, but see Muir et al. 1998). Neither island's constituents can be said to have a most recent common ancestor that is not also ancestral to some individuals from the other island. It is this same reasoning that makes the designation of Humans and Great Apes as separate groups phylogenetically nonsensical. The distribution of shared haplotypes is given in Figs.18, and 19. The most widespread haplotype, SHAGY, is found thoughout the contemporary range of orangutans. The WAF haplotype is found in Sabah and west Kalimantan. Figure 18 shows the geographic distribution of alleles that are shared for Cyt b.

Conclusion

There appears to be a fundamental difference in the distribution of mitochondrial haplotypes and nuclear alleles. Significant population structure within Sumatra and between some individuals in Sumatra and Borneo is evident in mitochondrial loci characterized to date (Xu and Arnason 1996, Zhi et al.. 1996). Nuclear loci, on the other hand, show little difference between populations while maintaining some heterogeneity overall. These differences in allele distribution may reflect the different home range requirements between male and female orangutans (Galdikas 1982). There is a higher probability of migration between demes among males than with females.

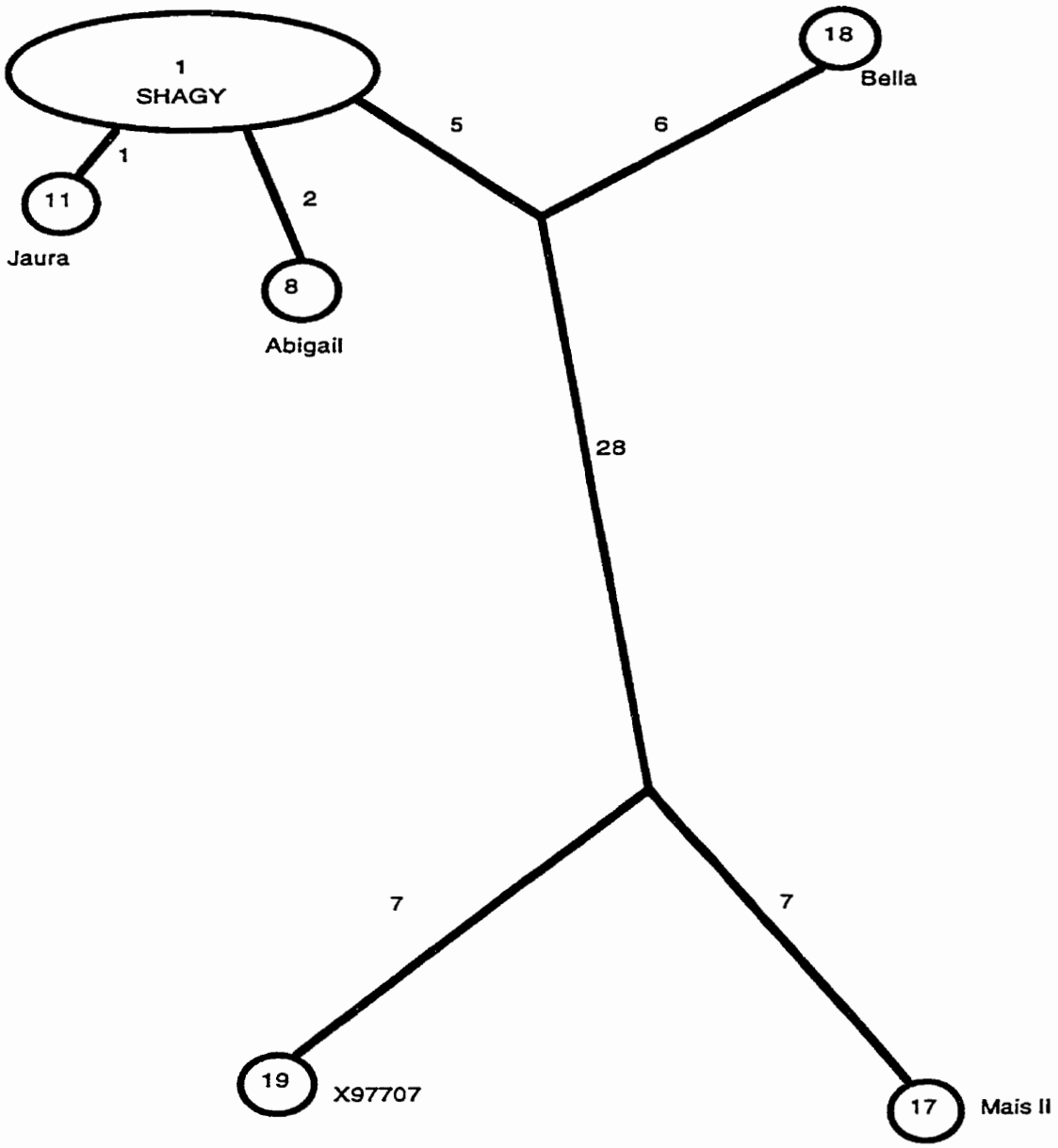


Figure 19: The cladogram of Sumatran Cyt b haplotypes. The number of pairwise differences between haplotypes are given on the bars between haplotypes.

Recent migration of males and females across the Borneo Sumatra land bridge in addition to relatively structured female, and panmictic male, gene flow may provide an explanation for present allele distribution.

Present population structure seems to reflect a high degree of gene flow at some time in the past. Contemporary human development has a dramatic effect on availability of routes between demes within Borneo and Sumatra into which ranges could expand. There is an immediate need to address how we engage this development so that we can mitigate the negative impact we have on the potential for the survival of the orangutan among other members of the biosphere.

Chapter 7
Survey of Other Loci

The preceding chapters support a model of population dynamics in which migration has played an important role in the distribution of alleles. The geographic distribution of mitochondrial haplotypes compared with that of nuclear alleles appears to be different. Mitochondrial genes are inherited maternally as a single locus. The distributions of ND3 and Cyt b haplotypes are consistent with genetic linkage. Furthermore, the haplotype diversity indicates recent, significant periods of genetic isolation. The distribution of haplotypes supports a paleodispersal model in which Sumatra was a migratory 'hub' between Borneo and the ancient distribution throughout South Asia.

The rate of evolution of the nuclear pseudo-gene is approximately 1/10 that of the mitochondrial rate. Because so few informative sites exist in the nuclear locus little can be said about the allele distribution. Geographically, gene flow appears to have been panmictic. All genetic types include individuals who were sampled widely over the entire range of Borneo and Sumatra. The difference in distribution of mitcohondrial and nuclear loci is influenced, in part, by differences in their rate of evolution. Another important influence on the different distributions is the likelihood for sex-specific migration. Male orangutans may utilize a much wider home range than females, a dynamic that would not be reflected in the analysis of a maternally inherited locus.

A brief survey of additional nuclear markers was undertaken to find an informative locus for population level inquiry. On a whole, the rate of evolution of nuclear loci is much slower, on the order of ten-fold, than mitochondrial genes. DNA sequence was determined for five nuclear non-coding regions including ψ Cyt b, which

was discussed in Chapter 6, introns four and six from arylsulfatase (A4, A6), intron three from lysozyme (L3), and a segment from the Y-specific SRY locus (SRY). In part because it is not clear what role homoplasy plays in their evolution, microsatellites were not included in this study. Since ψ Cyt b has been discussed (see Chapter 6), this chapter will be limited to a report on diversity in A4, A6, L3, SRY, and a short segment from the 5' end of COII including a spacer between COII and the tRNA for lysine. The Y chromosomal marker (SRY) was included in the analysis in an attempt to trace male specific migratory patterns.

Table 11: Names, origins, and haplotype designations for all individuals sampled for each locus. Gene designations are, arylsulfatase introns 4 (A4) and 6 (A6), lysozyme intron 3 (L3), Y specific marker (SRY), COII fragment (COII) and spacer between COII and the $tRNA^{lys}$ gene.

Name	Origin	A4	A6	L3	SRY	COII	Spacer
Bella	Sumatra	X	X	X	!	X	X
X97707*	Sumatra					X	X
Ruby	Sumatra	X				X	X
D38115*	Borneo					X	X
Rosemary	S.Central Kal.	X	X			X	X
Hobler Lily	S.Central Kal	X	X			X	
Davida	Borneo	X			!		
Tim	Sabah					X	
Supinah	W. Kalimantan	X					
Apollo Bob	W. Kalimantan	X	X		X		
Mellie	E. Kalimantan		X		!		
Julie	S.Central Kal	X	X		!		
Diane	Borneo		X				
Kelly	Borneo		X		!	X	
Patti	W. Kalimantan	X	X	X		X	
Roger	W. Kalimantan	X	X	X			X
Siswi	W. Kalimantan	X	X				
Mark	S.Central Kal	X	X	X	X		
Brook	S. Central Kal		X				
Herbie	W. Kalimantan	X	X	X	X		
Stan	S. Central Kal	X			X		
Doc	Borneo			X	X		
Abigail	Sumatra					X	X

! - Attempted PCR reaction but no product was detectable (all female individuals)

Methods

Table 11 gives a list of individuals analyzed for each of the gene regions. SRY sequence was determined for five males. Amplification was attempted for five females, but PCR products are not detectable for any of the female samples tested. Methodologies are outlined in Chapter 3.

TABLE 12: A comparison of loci. Maximum pairwise divergence between sequences are given, along with the transition/transversion ratio and indels for each gene.

	3'COII	SPACER	COII +SPACER	ARSA4	ARSA6	ARSA 4+6	LZM3	SRY
VARIATION	15/145= 10.3%	10/42= 23.8%	25/187= 13.4%	5/370= 1.35%	9/345= 2.61%	14/715= 1.96%	6/670= .896%	0
S/V	15/6	4/5 1 indel	19/11 1 indel	2/3	4/1 4 indels	6/4 4 indels	3/3	0
	Mitochondrial			Nuclear				

Results and Discussion

The sequence alignments for each of the loci studied in this chapter are given in appendices 7-11 while maximal divergences are reported in Table 12. The discrepancy in the rate of evolution between mitochondrial and nuclear loci reported in the last chapter persists in the loci reported here. It needs to be pointed out that the size of loci examined for this chapter are quite different; the mitochondrial loci together are only 187 bases long while the nuclear loci range in size from 345 (A4), to 868 (L3; Appendix 9). 868 base pairs of L3 were sequenced for two individuals (572 base pairs on average for the six individuals). Nevertheless, the maximum variation noted for the COII fragment (10.34%) is similar to that seen in Cyt b (8.7%) (Table 10) and ND3 (8.1%) (Table 7). There were insufficient variable sites in the COII fragment to be phylogenetically reliable. It would be worthwhile, in a future study, to sequence complete COII across a

wide sample of individuals from Borneo and Sumatra.

The nuclear loci sampled here ranged in variation from 0% (SRY), 0.90% (LZM), 1.35% (A4), to 2.61% (A6; see table 12). However, after correcting for size, the average variation across nuclear loci is 0.94%. This level of variation is similar to that found in ψ Cyt b (Chapter 6). It is interesting that the sequence alignment for L3 against two non-ape primates shows a high degree of similarity despite the loci being an intron which is supposedly not evolving under the constraint of evolution. The nuclear loci examined contain too few informative loci to be phylogenetically reliable. Nevertheless, clustering of individuals according to sequence identity, and accumulation of pairwise differences indicates maintenance of overall heterogeneity with respect to allele distribution. This difference in the distribution of mitochondrial and nuclear alleles might be expected if sex-specific migration is, or has been, significant.

Chapter 8
Conclusion

The relationship of the populations of orangutans on Borneo to those of Sumatra has, for some time, been in question (de Boer 1982, Xu and Arnason 1996, Jankowitz et al. 1990, Ryder and Chemnick 1993, Zhi et al. 1996, Muir *et al.* 1998). The debate has been chronicled as well in a number of popular articles that have appeared in, the NY Times, Science News, and Chronicle of Higher Education. Although varied, there is biogeographic consistency to the distribution of all alleles studied here. Nuclear alleles do not carry substantive phylogenetic information. Nevertheless, their distribution is consistent with panmictic gene flow since the alleles are not regionally partitioned.

The rate of evolution in mitochondrial genes is approximately ten times that of comparable nuclear loci. As a result, a limited opportunity for migration has existed for carriers of unique haplotypes. Haplotypes that are shared between small numbers of individuals in adjacent demes delineate possible migration routes, though not direction. The distribution of all mitochondrial haplotypes characterized for this study, and others, support hypothesized dispersal routes within the island of Borneo from East Kalimantan to West Kalimantan, and along the north western region. Sharing of haplotypes for a number of genes indicates a recent migration between Sumatra and deme(s) in south central and western Borneo.

The hypothesized model of dispersal includes the south Asian mainland as the original distribution. Migrations out of what is now Thailand, Vietnam, and Cambodia lead into Sumatra, Borneo and Java. Modeling of sea level changes suggests that land bridges appeared between these islands as a consequence of periodic glacial advances that occurred at least six times over the last million years (Chapter 2). Following the first

exodus from the mainland, the retreat of the glacial front resulted in the isolation of populations on Borneo, Sumatra, and the mainland. Significant portions of these populations must have remained genetically cut off from each other for an extended period, possibly until the last few hundred thousand years. Expansion of the orangutan population throughout Borneo would have been facilitated by the increase of coastal area during a glacial maximum. It appears that a recent migration has occurred from Borneo back into Sumatra. The directionality is implied from the closer similarity of some Sumatran haplotypes to Bornean types than to other Sumatrans. The sharing of haplotypes between the islands occurs for both mitochondrial (Chapters 4, 5, 6, and 7) and nuclear sequences (Chapters 6 and 7). This dispersal must have happened in the time since the initial isolation of populations.

Past migrations have an obvious impact on taxonomic interpretations. It is parsimonious to assume the most recent gene flow has followed species bifurcations. Although there are significant differences between species definitions, commonalties include the absence of interbreeding between distinct species. Individuals of a single species should be more genetically similar to each other than they are to members of a different species.

Recently, Xu and Arnason (1996) proposed that the orangutan of Borneo and Sumatra be assigned separate species status. The proposed taxonomic change is based on comparison of a single Sumatran mtDNA sequence to a single mtDNA sequence determined by Horai et al. (1982, GenBank accession #D38115). Two partial mtDNA sequences were also included in the analysis. An additional Sumatran D-loop sequence was determined but is identical to the D-Loop sequence found in the other Sumatran

orangutan. In the absence of any geographic information, I suspect the two were close relatives. In fact, the authors have no information about the geographic origins for any of their samples nor the Bornean individual (D38115).

The two sub-species are known to interbreed when brought together, so they clearly do not represent separate species under the Biological Species Concept (BSC) (Mayer 1965, Dobzhansky 1937). Xu and Arnason, however, have rejected the BSC, apparently in favor of the Genotypic Clustering (GC) definition (Mallet 1996). The GC definition requires the identification of *clusters* of individuals, in the absence of intermediates. I do not believe that the comparison of only two sequences (even complete ones) satisfies this definition. Nor do I believe that the addition of two partial sequences is sufficient to identify clusters or to demonstrate the absence of intermediates. The only basis for their conclusion appears to be the degree of divergence between the two complete mtDNA sequences. I will call this new species definition the mitochondrial DNA sequence divergence species definition (mtDNA $(SD)^2$).

Xu and Arnason's proposal is not the only proposal for separate species status for the orangutan populations. Jankowitz et al. (1991) alluded to this proposal, and Zhi et al. (1996), in a paper published concurrently with that of Xu and Arnason, also propose that the subspecies be elevated to species status based on 2-D gel data, mtDNA RFLP data, and an RFLP assessment of a single minisatellite locus.

Their conclusions appear untenable for a number of reasons. These include the choice of species definition and the reliance on a maternally inherited character for assessing gene flow. Since changes in the taxonomy of the orangutan are likely to have implications for their conservation, it is particularly important to not base these changes

on flawed arguments.

Species Definitions

Perhaps the greatest hurdle in finding an acceptable definition of a species has been our approach to the problem and its scope. Studies in natural history cover a vast range of phenomena. The process of speciation, limitations to reproductive interaction, and a plethora of life processes cannot be singularly typified for all organisms that make up our biosphere. Perhaps it is unrealistic to suppose a single definition of species can be constructed which will accommodate the set of all possible exceptions. Given these difficulties, it is tempting to make the definition quantifiable.

The mtDNA $(SD)^2$ defines a species based on mitochondrial (partial or complete) sequence divergence. Under this model, two populations are considered distinct species if their sequence divergence is equal to, or greater than, that of other accepted sibling species pairs. In addition to being slightly circular in logic, the mtDNA $(SD)^2$ makes a fundamental assumption: that the rate of speciation is proportional to the rate of acceptance of mutations (μ). This assumption requires that there is "a rate of speciation". For there to be a rate of speciation, speciation, as a process, must be of definable duration. Speciation is not a series of prescribed steps. The first assumption, that there is a fixed proportionality between μ and the process of speciation, assumes there are a definable number of mutations required for speciation. By this model two populations become separate species once they have achieved a certain level of sequence divergence. This notion rests on another term that lacks a foundation in reality: species level divergence. Jolly et al. (1995) pointed out that there is a 200 fold range of sequence

divergence in sibling vertebrate species (a range which would be decidedly bigger if it were more inclusive). I do not consider the mtDNA(SD)^2 to be a legitimate definition of a species since I find its underlying assumptions to be incorrect.

The GC species definition does not assume an algorithmically definable rate of speciation but does suffer from ambiguity and problems of applicability. Characterizing two populations with respect to each other would require complete genome sequence for each individual, a goal that is, at this time, untenable. Relying on molecular data to define the total gene flow between populations requires representative sampling in both the constituents of the populations and the segments of the genome. Defining a species after all, is not simply presenting a gene tree. It is not clear, under GC, how much DNA sequence is required (nor which locus is appropriate) to decide on clustering, nor is it clear how many individuals must be assessed. Finally, GC is really only applicable to overlapping populations (Mallet 1996) where reproductive isolation is real, and not simply a geographic barrier which may or may not persist.

Xu and Arnason do not, however, use the GC definition unless a single individual can be considered a cluster. Even the three Bornean individuals do not form a cluster. The differences among them, both in sequence and indels, would suggest that each represents a separate cluster (see fig. 1 Xu and Arnason 1996). Although the GC definition is a definite improvement on the mtDNA (SD)^2 , I do not think that sequence based definitions of a species are practicable solutions to taxonomic questions.

The Biological Species Concept (BSC) as presented by Mayr (1963) has received a great deal of criticism in that it is not directly testable in a large number of situations. This is, however, not a very sensible criticism in situations where it has been tested. Not

only do Bornean/Sumatran crosses result in viable offspring, but there are anecdotal reports that, at least in the captive setting, the orangutan prefers to "outcross". According to the Species Survival Plan (SSP) orangutan "Stud book", of the 304 captive orangutan in North American zoos, 88 are Bornean/Sumatran crosses (L. Perkins pers. comm.). I believe the BSC offers a good workable solution to defining species, certainly in the case of the orangutan.

Sampling

The sample set employed by Xu and Arnason is inadequate to answer any population genetics question let alone tackle a question of speciation. The authors use a single Sumatran mtDNA sequence. Obviously a single individual cannot represent an entire population (see previous chapters). To make matters worse, deBoer (1982) asserts there are at least two separate populations of orangutan in Sumatra, an assessment which is reflected by the molecular data. It is difficult to imagine how one individual could represent two populations. There is a similar problem with the Bornean samples. Part of the reason, apart from statistical significance, that it is important to have a number of samples from known geographic origins, is there appears to be an unusually high rate of acceptance of mutations in orangutan mtDNA (Ruvolo et al. 1994, Muir et al. 1994). There is no question that sequence divergence exists between individuals from the two islands but Xu and Arnason have made no attempt to assess the range of variation in the populations on these islands. The sampling employed by Xu and Arnason to fall short of what is necessary to draw any conclusion about population structure.

Character Choice

I do not want to suggest that there is an appropriate molecular character to compare as a sole criteria for taxonomic assessment. I believe sequence divergence to be an inadequate basis for defining a species. However, even if one holds the view that sequence divergence can answer taxonomic questions mtDNA cannot provide the necessary information. The mtDNA genome is inherited as a single genetic locus in a strictly maternal manner. A maternally inherited molecule is a poor choice for assessing gene flow in an organism with two sexes. The inability of mtDNA to provide information about gene flow in males is particularly problematic with the orangutan since male and female home ranges are often significantly different (Galdikas 1985 a,b).

Xu and Arnason present the complete mtDNA sequence of a Sumatran orangutan asserting that more sequence provides more information. Although the addition of more data seems at first compelling, there are diminishing returns in adding extra sequence, for a single locus, in any phylogenetic assessment. Zhi et al. (1996) recognize the need to include nuclear loci to understand gene flow in the orangutan. However their attempt to address this need by employing an RFLP analysis of a single minisatellite locus falls short. Given the high μ for minisatellites, in addition to increased likelihood for size homoplasy in populations that have been isolated for a significant period (Garza and Freimer 1996), I suggest minisatellites are not a good choice either.

Conservation Implications

A number of studies have pointed out that changing taxonomy has implications on conservation (Zhi et al. 1996, Mallet 1996, Uchida 1996). It is, therefore, important to not draw conclusions based on inadequate data. Part of the problem lies with the

translation of scientific conclusions for public consumption. Information is lost between the jargon of research journals and the lay-text of popular media. Here lies the danger in creating taxonomy that may have an influence on conservation strategy. It is a seductive argument that orangutan populations in Borneo and Sumatra should be maintained and conserved *because* they are separate species but by using this argument we create a vulnerability if an expanded date set argues against species level designation. Utilizing this argument also facilitates the prioritization of conservation efforts based on species status. It is essential that conservation issues be removed from the vagaries of academic systematics arguments. We cannot afford to risk the survival of endangered species while we attempt to answer a question we have been unable to answer for two centuries: the definition of a species. Are we basing global conservation and endangered species legislation on a term which we cannot define? In terms of conservation of the orangutan and other endangered species, a better more permanent case can be made for preserving diversity without tying the argument to ephemeral species definitions.

Discussion

In addition to a more philosophical argument against a species definition based on molecular divergence, the data presented in this thesis also argues against the lesser hypothesis that a simple bifurcation between the present populations which can be found on the islands of Sumatra and Borneo occurred and that a reproductive partition has persisted for the last million years over more than six major glacial epochs.

Data presented in this thesis do not support a bifurcation of orangutan species between the islands of Borneo and Sumatra. The common ground between species

definitions is the recognition that distinct species do not exchange genes. A number of mitochondrial and nuclear loci have been identified here for which individuals from Sumatra share more sequence similarity with Bornean orangutans than others from Sumatra. For all species definitions this observation fails the Bornean-Sumatran species split hypothesis. As a consequence of social structure, orangutan population structure is likely to be complicated by different male and female range size.

The orangutan of Borneo and Sumatra have been recognized as separate sub-species for over one hundred years. This distinction was made based on the geographic isolation which the populations face today and limited morphological data. While it is true that some Sumatran orangutans are morphologically distinct from some Bornean orangutans this distinction is not universal or perhaps even definitive. Furthermore, it is well established that many morphological features are subject to environmental and dietary influence. Cytological studies (Jankowitz et al. 1990, Ryder and Chemnick 1990) have shown a karyotypic landmark (chromosomal inversion) which has been used to identify Bornean and Sumatran orangutans in zoos. In Ryder and Chemnick's 1993 paper, however they point out that a number of the orangutans did not possess the karyotype predicted for their origin. These anomalies were explained by asserting that the orangutans had been misidentified. It may be that mislabeling is not the only explanation. Evidence from DNA sequence data from 13 mitochondrial and nuclear loci indicate that some Sumatran orangutans share more sequence similarity with Bornean orangutans than they do with their compatriots from Sumatra. Furthermore, there appears to be structure among Sumatran haplotypes which, even with a small sample set, seems to fall within three groupings of sequence similarity. For example, ND3 sequence data

indicate three distinct haplotypes among four individuals sampled. Each of the haplotypes is more than 6% divergent but Ruby and Abigail share sequence identity. The Cyt b fragment sequenced for this study provides similar phylogenetic inference: three or four distinct haplotypes. The overall picture in terms of population structure as discussed earlier appears to show some gene flow in males and females across Borneo and some migration across a recent land bridge opportunity to/ from Sumatra. It appears that at least three distinct populations have persisted for a longer period of time. This inter-island sequence identity renders even a subspecies split between the islands phylogenetically unsound; of course this view depends on the definition of a subspecies. It is important that we resist the urge to re-define concepts for the purpose of their convenient application. It may seem awkward to refer to "a currently reproductively isolated population(s)." Such a description perhaps tells us more of the information we want than the application of the term (sub-) species.

Future Directions

A better understanding of the population structure of the orangutan will require a more complete sampling of individuals from known origins. A notable limitation on this study has been the small number of Sumatran individuals sampled, particularly since all Sumatran individuals sampled are captive. Nevertheless, the extraordinary divergences apparent between individuals of Sumatran origin cannot fall with an increased sample set. A number of samples from wild individuals living in Sumatra will bolster what can be inferred about Sumatran population structure especially since tremendous genetic diversity among Sumatran orangutans is apparent.

It is not surprising that different segments of a single locus, the mitochondrial genome, imply the same population structure. Since the mitochondrial genome is inherited maternally as a non-recombining locus one expects it act in unison phylogenetically. It would add to the understanding of these populations to examine a number of different microsatellite loci keeping in mind the possibility of homoplasmy, which may grow with increased time since isolation. A Y-specific marker, which is phylogenetically informative at the population level, may provide insight into the migratory behaviors of males. It would also be very interesting to conduct a large-scale paternity analysis across many populations. Broad scale population analysis leads to a synthesis of data into phylogeographic appraisal.

In the future I would very much like to participate in a study using sub-fossil tissue remains from China, Vietnam, Peninsular Malaysia, Sumatra, Java, and Borneo for possible DNA extraction. These tissue remains would be a small sampling from a much larger, though extinct, population of orangutans that existed as recently as the beginning of the most recent glacial retreat. Genetic diversity of this historic population would make for an interesting comparison to contemporary population structure. A comparison of this nature may give rare insight into a dynamic in population size and distribution over a longer period of time than has been previously undertaken.

The orangutan is undergoing a severe population decline. There are many regions within their range where the orangutan is no longer found. These demographics have important effects on the population genetics of the orangutan that may affect their survival. A more complete understanding of conservation genetics requires an ecological context. Therefore, extending a phylogeographic appraisal to a multi-species level would

be an informative long-term study.

Appendix 1: List of Individuals included in analysis.

Name	Origin	Source	Tissue
Bella	Sumatra	Toronto Zoo	blood & hair
X97707*	Sumatra	genbank	
Ruby	Sumatra	Dallas Zoo	hair
Abigail	Sumatra	Toronto Zoo	hair
D38115*	Borneo	genbank	
Rosemary	S.Central Kal.	BMF Galdikas	blood
Maggie	Bintulu	collected by CM	hair
Supinah	W. Kalimantan	BMF Galdikas	blood
Hobler Lily	S.Central Kal	"	"
Davida	Borneo	"	"
Apollo Bob	W. Kalimantan	"	"
CLO248	Borneo	"	"
Mellie	E. Kalimantan	"	"
Julie	S.Central Kal	"	"
Diane	Borneo	"	"
Kelly	Borneo	"	"
Patti	W. Kalimantan	"	"
Roger	W. Kalimantan	"	"
Siswi	W. Kalimantan	"	"
Mark	S.Central Kal	"	"
Brook	S. Central Kal	"	"
Herbie	W. Kalimantan	"	"
Stan	S. Central Kal	"	"
Doc	Borneo	"	"
CLO1239	Borneo	"	"
CLO1228	Borneo	"	"
ROBOT	Sabah	collected by CM	hair
Semenduh	Sabah	"	"
H2	Sabah	"	"
V03	Sabah	"	"
V04	Sabah	"	"
X01	Sabah	"	"
W03	Sabah	"	"
W05	Sabah	"	"
W02	Sabah	"	"
Tim	Sabah	"	" (and faeces)
Gensusuli	Sabah	"	"
Kim Long	Sabah	"	"
Baboon	W. Kalimantan	"	"
Bebeta	W. Kalimantan	"	"
Lemot	Lubok Ntu	"	"
WO7	Sabah	"	"
AH FONG	Sabah	"	"
RANTO	Borneo	"	"
TEROGON	Sabah	"	"
RIA	Sabah	"	"
GENTUTU	W. Kalimantan	"	"
Y01	Sabah	"	"
Meliwar	Sabah	"	"
Gamgar	Sumatra		
JaraKong	W. Kal	"	"
Jdura	Sumatra	Toronto Zoo	"
Anna(X97717)*	Borneo	Genbank	?
U38269*	Borneo	"	?

*= GenBank acession number

Appendix 2: Field Trip Checklist.

Contact and Permits:
- establish a contact/ host person
- CITES (United Nations Convention on International Trade in Endangered Species) if sample is CITES controlled
- work/ entry Visa

Health
- innoculation (eg. hepatitus, yellow fever)
- Anti-Malaria pills (many side effects)
- sunscreen

Oher
- DEET
- proper clothing/ footware for weather
- mosquito netting
- mosquito coils
- good back pack
- ceramic water filter
- first-aid kit
- laundry soap and scrub brush
- good durable camera
- note book, pens, labels, markers
- **plastic bags, water proof tubes with caps, 70%EtOH**

Appendix 3: PCR primers.

Name	Sequence	Comments
arg/ND3/gly		
L9415	(5' CCATCTACTGATGAGGGTCTTAC)	Designed by CM were used to amplify ND3, arg and gly.
H9975	(5' ATTAGGTGTGAGCGGTAGAC))	
COII/lys		
8124	5'GACGTCTAAACCAAACCACT	
8366	5'GTAGTATTTAGTTGGGGCAT	(CM) designed to amplify lys,
Dloop/pro		
L15323	5'AATGAACCTGCCCCTGTAGT	
H17	5'CAGTGTATTGCTTTGGATGG	(CM) designed to amplify pro.
Pseudo-Cyt. b:		
PCb 1 (L15415)	5'-GACAAAATCACCTTCCACCCTTAC	
Pcb 2 (H15835)	5'-CGGTGAGTGGTATGAGGGCG	(Collura and Stewart 1995):
Cyt. b:		
ECB1	5'CACGAAACAGGATCAAATAACCC	
Ecb2	5'ATTTTCAGGTTACAAGGCTGGCG	(Collura and Stewart 1995)
Arylsulfatase Intron 4		
ARSA4F:	5' GGGGACCTGGGGCTGCTTGAA	
ARSA4R:	5' CCAAGGCAGGCTCTCGGACAC)	Melting temp~63C designed by CM from published exon sequence
Arylsulfatase Intron 6		
ARSA6F:	5' GCCCCACTGCCCAATGTCACC	
ARSA6R:	5' CCAGTCCGCACAGCAAAAACC	Melting temp ~63Cdesigned byCM from published exon sequence
LYSOZYME Intro 3		
LZM3F (or WM15)	5' CTTTGCTGCAAGATAACATC	
LZM3R (or WM2)	5' ACCTTTCACTTAATTCCTACTCCC	(melting temperature 55C (ref: MEDLINE 97144349: (Messier and Stewart 1997));
LZM3Fb:	5' CCTCCCGGGTTCAAGCAAGTCTC,	LZM nested primer melting tempeture~60C (design by CM from LZM determined sequence)
LZM3Rb	5' GGCAAAACCCCATCTCTACTAAA)	
Y Specific		
SRY 1:	5' CTTGAGATTGAATACATTGTCAGGG:	
SRY 2	5' AGGTCTTTGTAGCCAATGTTACCCG	melting temperature ~ 60C (ref: Whitfield et al. 1993)

Appendix 4: ND3 sequence alignment.

```
GRRMaHl       ATT AAC TTC GTC CTA GCT CTG ACA ATT AAC ACC CTC CTT GCC CTA CTA CTA
MELLIE        ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
"BABOON"      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
AB248         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
JULIE         ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ...
DIANE         ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ...
KELLY         ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ...
PATTI         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DAVLEE        ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ROGER         ... ... ... ... ... ... ..A ... ... ... ... T.. ... ... ... ... ...
SISWI         ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ...
MARK          ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BROOK         ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ...
HERBIE07      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
STAN07        ... ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DOC07         ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ...
BEBETA        ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ...
LEMOT         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BELLA         ... ... ... ... ... ... ..A ... ... ... ... ... ..C ... ... ... ...
CLO239        ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SUMND3        ... ... ... ... ... ... ..A ... G.C ... ... ... ..A ... ..G ... ...
WO7           ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SHVVXYWWTGB   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
                                                                              102
GRRMaHl       ATA ATC CTC ACA TTC TGA CTA CCC CAA CTT AAC CCC TAC ATA GAA AAA TCC
MELLIE        ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ...
"BABOON"      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
AB248         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
JULIE         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DIANE         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
KELLY         ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ...
PATTI         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DAVLEE        ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ROGER         ... ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ...
SISWI         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
MARK          ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BROOK         ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ...
HERBIE07      ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ...
STAN07        ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DOC07         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BEBETA        ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ...
LEMOT         ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ...
BELLA         ... .C. A.. ... ... ... ... ... ... ... ... ... ... ... ... ... ...
CLO239        ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SUMND3        ... .C. A.. ... ... ... ... ..A ... ..C T.. ... ... ... ... ... ...
WO7           ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ...
SHVVXYWWTGB   ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ...
```

																	153
GRRMaH1	GAC	CCA	TAC	GAA	TGT	GGA	TTT	GAC	CCC	GCA	TAC	CCT	GCC	CGC	ATT	CCT	TTC
MELLIE	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
"BABOON"	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
AB248	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
JULIE	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
DIANE	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
KELLY	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
PATTI	...	...	...	...	..A	...	...	...	...	...	...	...	...	...	...	...	...
DAVLEE	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ROGER	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
SISWI	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
MARK	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BROOK	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
HERBIE07	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
STAN07	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
DOC07	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BEBETA	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
LEMOT	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BELLA	...	...	..T	...	...	...	...	...	..T	...	...	...	...	...	..C	...	...
CL0239	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
SUMND3	...	...	...	...	...	...	...	...	...	...	...	..C	..T	...	...	...	...
WO7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
SHVVXYWWTGB	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

																	204
GRRMaH1	TCC	ATA	AAA	TTC	TTC	CTA	GTA	GCC	ATC	ACC	TTC	CTA	CTA	TTC	GAT	CTA	GAA
MELLIE	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
"BABOON"	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
AB248	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
JULIE	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
DIANE	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
KELLY	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
PATTI	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
DAVLEE	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ROGER	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
SISWI	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
MARK	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BROOK	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
HERBIE07	...	...	...	...	...	...	...	...	...	...	...	...	..G	...	...	...	...
STAN07	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
DOC07	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BEBETA	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
LEMOT	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BELLA	...	...	...	...	...	...	...	...	...	...	...	...	..G	...	..C	...	...
CL0239	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
SUMND3	...	...	...	..T	...	T..	...	...	...	...	...	...	...	...	..C	...	...
WO7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
SHVVXYWWTGB	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

```
                                                                       255
GRRMaH1     ATT GCC CTA CTA CTA CCC CTA CCA TGA GCC CTA CAA ACA ACC AAC TTA CCA
MELLIE      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
"BABOON"    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
AB248       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
JULIE       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DIANE       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
KELLY       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
PATTI       ... ... ... ... ... ... ... ... ... .G. ... ... ... ... ... ... ...
DAVLEE      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ROGER       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SISWI       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
MARK        ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BROOK       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HERBIE07    ... ... ... ... ... ... ... ... ... .G. ... ... ... ... ... ... ...
STAN07      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T
DOC07       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BEBETA      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
LEMOT       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BELLA       ... ... ... ... ... ... ... ... ..G ... ... ... ... G.. ... ... ...
CLO239      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SUMND3      ..C ... ..G ... ... ... ..G ... ..G ... ... ... ... ... ... ... ...
WO7         ... ... ... ... ... ... ... ... ... .G. ... ... ... ... ... ... ...
SHVVXYWWTGB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
                                                                       306
GRRMaH1     CTA ATA ACC ACA TCA TCG CTT ATA TTA ATT ATT ATC CTG GCC CTG GGC CTA
MELLIE      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
"BABOON"    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
AB248       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
JULIE       ... ... ... G.. ... ... ... ... ... ... ... ... ..A ... ... ... ...
DIANE       ... ... ... G.. ... ... ... ... ... ... ... ... ... ... ... ... ...
KELLY       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
PATTI       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DAVLEE      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ROGER       ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ..A ... ...
SISWI       ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ..A ... ...
MARK        ... ... ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ...
BROOK       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HERBIE07    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
STAN07      .A. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DOC07       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BEBETA      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
LEMOT       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BELLA       ... ... ... ..G ... ... ... ... ... ... ..C ... ..A ... ..A ... ...
CLO239      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ...
SUMND3      ... ... ..T ... ... ..A ... ... ... ... ..C ... ..A ... ..A ... ...
WO7         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SHVVXYWWTGB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
```

													345
GRRMaH1	ACT	TAC	GAA	TGA	TCC	CAA	AAA	GGA	TTA	GAC	TGA	ACC	GAA
MELLIE	...	...	...	...	...	...	...	...	...	..A	...	...	...
"BABOON"	..C	...	...	...	...	...	...	...	...	...	...	...	...
AB248	...	...	...	...	...	...	...	...	...	..A	...	...	...
JULIE	...	...	...	...	...	...	...	...	...	...	...	...	...
DIANE	...	...	...	...	...	...	...	...	...	...	...	...	...
KELLY	...	...	...	...	...	...	...	...	...	...	...	...	...
PATTI	...	...	...	...	...	...	...	...	...	...	...	...	...
DAVLEE	...	...	...	...	...	...	...	...	...	...	...	G..	...
ROGER	...	...	...	...	...	...	...	...	...	...	...	...	...
SISWI	..C	...	...	...	...	...	...	...	...	...	...	...	...
MARK	...	...	...	...	...	...	...	...	...	...	...	...	...
BROOK	...	...	...	...	...	...	...	...	...	...	...	G..	...
HERBIE07	...	...	...	...	...	...	...	...	...	...	...	G..	...
STAN07	...	...	...	...	...	...	...	...	...	...	...	G..	...
DOC07	..C	...	...	...	...	...	...	...	...	...	...	...	...
BEBETA	...	...	...	...	...	...	...	...	...	...	...	G..	...
LEMOT	..C	...	...	...	...	...	...	...	...	...	...	...	...
BELLA	...	...	..G	...	..A	...	...	...	...	..T	...	G..	...
CLO239	...	...	...	...	...	...	...	...	...	...	...	...	...
SUMND3	...	...	...	...	..A	...	...	...	...	...	...	G..	...
WO7	...	...	...	...	...	...	...	...	...	...	...	G..	...
SHVVXYWWTGB	...	...	...	...	...	...	...	...	...	...	...	G..	...

Appendix 5: Cyt b sequence alignment.

```
                                                                              54
WAF      ACT CCG ATA AAA TCA CTT TCC ACC CCT ACT ATA CAA TTA AAG ACA TCC TAG GCC
SHAGY    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ANNA     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
TIM      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
GENEBNK  .T. ... .C. ... ... .C. ... ... ... ... .CT ... .C. ... .T. ... ... ...
MAGGIE   ... ... ... ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ...
ABIGAIL  ... ... ... ... ... ... ... ... ... ... .C. ... ... ... ... ... ... ...
KIMLONG  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BEBETA   .T. .T. .C. ... ... .C. ... ... .T. ... ... ... .C. ... .T. ... ... ...
Jdura    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DOC      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
H5       .?. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
MAWAR    .T. ... .C. ... ... .C. ... ... ... ... .C. ... .C. ... ... ... ... ...
BELLA    ... ... .C. ... ..G .C. ... ... ... ... .C. ... ... ... ... ... ... ...
SUMCYTB  .T. ... .C. ... ... .C. ... ... ... ... .C. ... .C. ... ... ... ... ...
HSACYTB  .T. ... ... ... ... .C. ... ... .T. ... .C. ... .C. ... ... C.. .C. ..T
HSACYTB  .T. ... ... ... ... .C. ... ... .T. ... .C. ... .C. ... ... C.. .C. ..T

                                                                             108
WAF      TAC TCC TTT TTC TCC TCG CCC TAA TAA CAT TAA CAC TAC TCT CAC CAG ACC TCC
SHAGY    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ANNA     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
TIM      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
GENEBNK  ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ...
MAGGIE   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ABIGAIL  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
KIMLONG  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BEBETA   ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Jdura    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DOC      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
H5       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
MAWAR    ... ... ... ... ... ... ... ... C.. ..C ... ... ... ... ... ... ... ...
BELLA    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .T. ...
SUMCYTB  ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ...
HSACYTB  ... .T. .C. .C. .T. ..T ..T ... .G. ... ... ... ..T ... ... ... ... ...
HSACYTB  ... .T. .C. .C. .T. ..T ..T ... .G. ... ... ... ..T ... ... ... ... ...

                                                                          162
WAF      TAA GCG ACC CAG ACA ACT ACA CCT TAG CTA ACC CCC TAA GCA CCC CAC CCC ACA
SHAGY    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ANNA     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
TIM      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
GENEBNK  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
MAGGIE   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ABIGAIL  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
KIMLONG  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BEBETA   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Jdura    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DOC      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
H5       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
MAWAR    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BELLA    ... ... ... ... ... ... ... ... ... .C. ... ... ... ... ... ... ... ...
SUMCYTB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HSACYTB  ..G ... ... ... ... .T. .T. ..C ... .C. ... ..T ... A.. ... .T. ... ...
HSACYTB  ..G ... ... ... ... .T. .T. ..C ... .C. ... ..T ... A.. ... .T. ... ...
```

216

WAF	TTA	AGC	CCG	AAT	GAT	ACT	TCC	TAT	TCG	CCT	ACG	CAA	TCC	TAC	GAT	CCG	TCC	CCA
SHAGY	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ANNA	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
TIM	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
GENEBNK	...	.A.	...	.G.	...	.T.	...	...	...	...	...	...	...	...	...	...	...	...
MAGGIE	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ABIGAIL	...	.?.	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
KIMLONG	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BEBETA	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Jdura	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
DOC	...	...	A..	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
H5	...	...	A..	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
MAWAR	...	.A.	...	.G.	...	.T.	...	...	...	...	...	...	...	...	...	...	...	...
BELLA	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
SUMCYTB	...	.A.	...	...	...	.T.	...	...	...	...	...	...	...	...	...	...	...	...
HSACYTB	.C.	...	...	...	...	.T.	...	...	...	...	..A	...	.T.	.C.	...	...	...	.T.
HSACYTB	.C.	...	...	...	...	.T.	...	...	...	...	..A	...	.T.	.C.	...	...	...	.T.

270

WAF	ACA	AAC	TAG	GAG	GTG	TAA	TAG	CCC	TCA	TGC	TAT	CTA	TCC	TAA	TCC	TAA	CAA	CAA
SHAGY	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ANNA	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
TIM	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
GENEBNK	...	...	...	...	...	.G.	...	...	...	...	...	.C.	...	...	...	...	...	...
MAGGIE	...	...	...	...	...	.??	...	...	...	...	...	.?.	...	...	...	...	...	...
ABIGAIL	...	...	...	...	...	...	...	...	...	...	...	.C.	...	...	...	...	...	...
KIMLONG	...	...	...	...	...	..T	...	G..	...	...	...	...	...	...	...	.T.	...	...
BEBETA	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Jdura	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
DOC	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
H5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
MAWAR	...	...	...	...	...	...	...	...	...	...	...	.C.	...	...	...	...	..G	...
BELLA	.T.	.G.	...	...	.C.	...	...	...	...	...	...	.C.	...	...	...	...	...	...
SUMCYTB	...	...	...	...	...	...	...	...	...	.A.	...	.C.	...	...	...	...	...	...
HSACYTB	...	...	...	...	.C.	.CC	.T.	...	.AT	.A.	...	.C.	...	.C.	...	..G	...	T..
HSACYTB	...	...	...	...	.C.	.CC	.T.	...	.AT	.A.	...	.C.	...	.C.	...	..G	...	T..

324

WAF	TCC	CCG	CTC	TCC	ACA	CAT	CCA	AGC	AAC	AGA	GCA	TAA	CAT	TCC	GCC	CAT	TAA	GCC
SHAGY	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ANNA	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
TIM	...	...	...	...	...	...	...	...	...	...	...	...	...	.T.	...	...	...	...
GENEBNK	...	.T.	.C.	.T.	...	T..	...	...	...	...	...	...	...	.T.	...	...	.G.	...
MAGGIE	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ABIGAIL	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
KIMLONG	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BEBETA	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Jdura	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
DOC	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
H5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
MAWAR	...	.T.	.C.	.T.	...	T..	...	...	...	...	...	...	...	.T.	...	..C	.G.	...
BELLA	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
SUMCYTB	...	.T.	.C.	.T.	...	TG.	...	...	...	...	...	...	...	.T.	...	...	.G.	...
HSACYTB	...	..A	TC.	...	.T.	T..	...	.A.	...	.A.	...	...	T..	.T.	...	..C	...	...
HSACYTB	...	..A	TC.	...	.T.	T..	...	.A.	...	.A.	...	...	T..	.T.	...	..C	...	...

```
                                                                          378
WAF      AAT TCC TAT ATT GAC TCT TAA TCA CCG ACC TCC TAG TTC TCA CCT GAA TTG GAG
SHAGY    ... ... ... .C. ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ANNA     ... ... ... .C. ... ... ... ... ... ... ... ... ... ... ... ... ... ...
TIM      ... ... ... .C. ... ... ... ... ... ... ... ... ... ... ... ... ... ...
GENEBNK  ... ... ... ... ... .T. ... ..G ... ... .T. ..A ... ... ... ... ... ...
MAGGIE   ... ... ... .C. ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ABIGAIL  ... ... ... .C. ... ... ... ... ... ... ... ... ... ... ... ... ... ...
KIMLONG  ... ... ... .C. ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BEBETA   ... ... ... .C. ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Jdura    ... ... ... .C. ... ... ... ... ... ... ... A.. ... ... ... ... ... ...
DOC      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
H5       ... ... ... .C. ... ... ... ... ... ... ... ... ... ... ... ... ... ...
MAWAR    ... ... ... ... ... .T. ... ..G T.. ... .T. ..A ... ... ... ... .C. ...
BELLA    ... ... ... .C. ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SUMCYTB  ... ... ... ... ... .T. ... ..G ... ... .T. ..A ... ... ... ... ... ...
HSACYTB  ... CA. .T. ... ... ..C ..G C.G .A. ... ... .CA ... .A. ... ... .C. ...
HSACYTB  ... CA. .T. ... ... ..C ..G C.G .A. ... ... .CA ... .A. ... ... .C. ...

                                                                          432
WAF      GAC AAC CAG TAA GCT ACC CCT TCA TTA CTA TTG GCC AAG TAG CAT CCG TAC TAT
SHAGY    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ANNA     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
TIM      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
GENEBNK  ... ... ... ... ... .T. ... ... .C. .C. ..A ... ... ... ... ..A C.T ...
MAGGIE   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ABIGAIL  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
KIMLONG  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BEBETA   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Jdura    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DOC      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
H5       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
MAWAR    ... ... ... ... ... .T. ... ... .C. .C. ..A ... ... ... ... ..A C.T ...
BELLA    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SUMCYTB  .G. ... ... ... ... ... ... ... .C. .C. ..A ... ... ... ... ..A C.T .G.
HSACYTB  ... ... ... ... ... ... .T. .T. CC. TC. ... .A. ... ... ... ... ... ...
HSACYTB  ... ... ... ... ... ... .T. .T. CC. TC. ... .A. ... ... ... ... ... ...

                                                                          486
WAF      ACT TTA CCA CTA TCC TAC TCC TTA TAC CAA CCT CTT CCC TGA TCG AAA ACT ACA
SHAGY    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ANNA     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
TIM      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
GENEBNK  ... .C. .T. ... ... .T. .A. ... ... ..G ... T.. ... .A. ... ... ..C ...
MAGGIE   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
ABIGAIL  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
KIMLONG  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
BEBETA   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Jdura    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DOC      ... ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ...
H5       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
MAWAR    ... .C. .T. T.. ... .T. .A. ... ... ..G ... .C. ... ... ... ... ... ...
BELLA    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SUMCYTB  ... .C. .T. ... ... .T. .A. ... ... ..G ... ... ... ... ... ... ..C ...
HSACYTB  ... .C. .A. .A. ... ..A ... .A. ... ... .TA TC. ... .A. .T. ... ..A .A.
HSACYTB  ... .C. .A. .A. ... ..A ... .A. ... ... .TA TC. ... .A. .T. ... ..A .A.
```

```
            492
WAF      TAC TCA
SHAGY    ... ...
ANNA     ... ...
TIM      ... ...
GENEBNK  ... ...
MAGGIE   ... ...
ABIGAIL  ... ...
KIMLONG  ... ...
BEBETA   ... ...
Jdura    ... ...
DOC      ... ...
H5       ... ...
MAWAR    ... ...
BELLA    ... ...
SUMCYTB  ... ...
HSACYTB  ... ...
HSACYTB  ... ...
```

Appendix 6: ψ Cyt b sequence alignment.

```
                                                                          60
BAJ         TACACAACCA AAGATATTCT AGGTTTAATT TTTCTCCTCC TCCTTTTAAT AATTCTAGTA
DIANE       .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
APOLLOBOB   .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... ..C....... .....-.... .......... .......... ..........
SHERB       .......... .......... .......... .......... .......... ..........
PPY         .......... .-.....-.. .....-.... ..-....... .......... ..........

                                                                         120
BAJ         CTGTTTTCGC CTGACCTCCT GGGTGACCCA GATCATTACA CCTTAGTCAA CCCCCTAAAT
DIANE       .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
APOLLOBOB   .......... .......... .......... .-........ .......... ..........
HOBLERLILY  .......... .......... .......... .......... .......... ..........
SHERB       .......... .......... .......... .......... .......... ..........
PPY         .......... .......... .......... .......... .......... ..........

                                                                         180
BAJ         ACCCTGCCCC ACACTAAACC ACAGTGATAC TTTTTATTTG CATATGCAAT CTTATGATCC
DIANE       .......... .......... .....T.... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
APOLLOBOB   .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... .......... .......... .......... .......... ..........
SHERB       .......... .......... .......... .......... .......... ..........
PPY         .......... .......... .......... .......... .......... ..........

                                                                         240
BAJ         ATCCCCAATA AATTAGGAGG CGTACTGGCC CTTCTATTAT CCATTCTCAT TCTAGCAGTT
DIANE       .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
APOLLOBOB   .......... .......... .......... .......... ......A... ..........
HOBLERLILY  .......... .C........ .......... .......... .......... ..........
SHERB       .......... .......... .......... .......... .......... ..........
PPY         .......... .......... .......... .......... .......... ..........

                                                                         300
BAJ         ATTCCTGCAC TCCACACATC CAAACAACAA AGCATCATAT TTCACCCATT AAGTCAGTAT
DIANE       .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
APOLLOBOB   .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... ....T.T... .......... .......... .......... ..........
SHERB       .......... .......... .......... .......... .......... ..........
PPY         .......... .......... .......... .......... .......... ..........
```

```
                                                                      360
BAJ         CTGTTCTGAA TCTTAGTCAC CGACCTATTC ACACTCACAT GAATCAGAGG ACAGCCAGTT
DIANE       .......... .......... .......... .......... .......... ..........
TIM         .......... .......... .......... .......... .......... ..........
APOLLOBOB   .......... .......... .......... .......... .......... ..........
HOBLERLILY  .......... .......... .......... .......... .......... ..........
SHERB       .......... .......... .......... .......... .......... ..........
PPY         .......... .......... .......... .......... .......... ..........


                                                                      419
BAJ         GAACAGCCTT TTATTACCAT TGGACAGATA GCATCTA?GA TATATTTCTC TATTATTCT
DIANE       .......... .......... .......... .......... .......... .........
TIM         .......... .......... .......... .......... .......... .....C...
APOLLOBOB   .......... .......... ........G. .......... .......... .........
HOBLERLILY  .......... .......... .......... .......... .......... .........
SHERB       .......... .......... .......... ........A.. .......... .........
PPY         .......... .......... .......... ........A.. .......... .........
```

Appendix 7: Arylsulfatase Intron 4 sequence alignment.

```
                                                                      60
HUMAN      CGCTGGTCAT CTTCACTGCA GACAATGGGT ATGCCAGCAG GGCAGCTGGG TGCTCCGGCC
ROGER      .......... .......... .....C.... .......... ......A... ..........
SISWI      .......... .......... .....C.... .......... ......A... ..........
BELLA      ???????... .......... .....C.... .......... ......A... ..........
PATTI      .......... .......... .....C.... .......... ......A... ..........
APOLLO BOB .......... .......... .....C.... .......... ......A... ..........
HERBIE     .......... .......... .....C.... .......... ......A... ..........
HOBLERLILY ?????????? ?......... .....C.... .......... ......A... ..........

                                                                     120
HUMAN      CTGTCACGGG CCAGGGCCTG GAGGCCTTGC AGTTCAGCTG CTTGCCAAGA ACATAGTGGG
ROGER      .......A.. .......... .G........ .......... .......... .......TA.
SISWI      .......A.. .......... .G........ .......... .......... ..........
BELLA      .......A.. .......... .G........ .......... .......... ..........
PATTI      .......A.. .......... .G........ .......... .......... ..........
APOLLO BOB .......A.. .......... .G........ .......... .......... ..........
HERBIE     .......A.. .......... .G........ .......... .......... ..........
HOBLERLILY .......A.. .......... .G........ .......... .......... ..........

                                                                     180
HUMAN      TGAGGGGGTG CCAGGAGATG CTGGCCACGT TGCAGGGGCC CAAGGTGTAG TCAGGAGACA
ROGER      .......... T......... .......... .......A.. .......... ..........
SISWI      .......... T......... .......... .......A.. .......... ..........
BELLA      .......... T......... .......... .......A.. .......... ..........
PATTI      .......... T......... .......... .......A.. .......... ..........
APOLLO BOB .......... T......... .......... .......A.. .......... .AT.......
HERBIE     .......... T......... .......... .......A.. .......... ..........
HOBLERLILY .......... T......... .......... .......A.. .......... ..........

                                                                     240
HUMAN      CAGGTGCACA GAGAGCTGGT CTTGGTAGGC CTGGGAGGTG CCGGGCTCAT GCTGGGCACC
ROGER      .......... .......... .......... .......A.. ........G. ..........
SISWI      .......... .......... ........A. .......A.. ........G. ..........
BELLA      .......... .......... .......... .......A.. ........G. ..........
PATTI      .......... .......... .......... .......A.. ........G. ..........
APOLLO BOB .......... .......... .......... .......A.. ........G. ..........
HERBIE     .......... .......... .......... .......A.. ........G. ..........
HOBLERLILY .......... .......... .......... .......A.. ........G. ..........

                                                                     300
HUMAN      TCCGGGCAAG CTTTGTGACT TAGAGGTGTG GGGCCACTGG TCACCCTCGG TGGCTCAGAG
ROGER      .......... .......... .....C.... .....G.... .......... ..........
SISWI      .......... .......... .....C.... .....G.... .......... ..........
BELLA      .......... .......... .....C.... .....G.... .......... ..........
PATTI      .......... .......... .....C.... .....G.... .......... ..........
APOLLO BOB .......... .......... .....C.... .....G.... .......... ..........
HERBIE     .......... .......... .....C.... .....G.... .......... ..........
HOBLERLILY .......... .......... .....C.... .....G.... .......... ..........
```

```
                                                                          360
HUMAN       GCTGTGGCTC CATGGCTCAT GAGCGCCTCC TGTGTCCCAG ACCTGAGACC ATGCGTATGT
ROGER       .......... .......T.. ....T..... .......... .......... ..........
SISWI       .......... .......T.. ....T..... .......... .......... ..........
BELLA       .......... .......T.. ....T..... .......... .......... ..........
PATTI       .......... .......T.. ....T..... .......... .......... ..........
APOLLO BOB  .......... .......T.. ....T..... .......... .......... ..........
HERBIE      .......... .......T.. ....T..... .......... .......... ..........
HOBLERLILY  .......... .......T.. ....T..... .......... .......... ..........

                   370
HUMAN       CCCGAGGCGG
ROGER       ..........
SISWI       ..........
BELLA       ..........
PATTI       ..........
APOLLO BOB  ..........
HERBIE      ..........
HOBLERLILY  ..........
```

Appendix 8: Arylsulfatase Intron 6 sequence alignment.

```
                                                                            60
HUMAN       GGATGGCTTT GACCTCAGCC CCCTGCTGCT GGGCACAGGC AAGGTAGGGC CGGTGACCCC
PATTI       .......... .......... .......... .......... .......... ..........
BROOK       ?????????? ?......... .......... .......... ..-------. ..........
ROGER       ???....... .......... .......... .......... .......... ..........
SISWI       ???????... .......... .......... .......... .......... ..........
ROSEMARY    .......... .......... .......... .......... .......... ..........
1228        ?????????. .......... .......... .......... .......... ..........
DOC         .......... .......... .......... .......... .......... ..........
BELLA       .......... .......... .......... .......... .......... ..........
APOLLO BOB  .......... .......... .......... .......... .......... ..........
HERBIE      ???????... .......... .......... .......... .......... ..........
HOBLERLILY  ?????????? ?......... .......... .......... .......... ..........

                                                                           120
HUMAN       TGATCCCAGA TCCTTGGCCC CTGTCCTGGC CTTCCCCTGG GGTGAGTGTG G-CAGTGC-T
PATTI       .......... .......... ....T..... .C........ .......... .G....T.C.
BROOK       .......... .......... ....T..... .C........ .......... .G....T.C.
ROGER       .......... .......... ....T..... .C........ .......... .G....T.C.
SISWI       .......... .......... ....T..... .C........ .......... .G....T.C.
ROSEMARY    .......... .......... ....T..... .C........ .......... .G....T.C.
1228        .......... .......... ....?..... .C........ .......... .G....T.C.
DOC         .......... .......... .......... .C........ .-........ .G....T.C.
BELLA       .......... .......... .......... .C........ .......... .G....T.C.
APOLLO BOB  .......... .......... ....T..... .C........ .......... .G....T.C.
HERBIE      .......... .......... ....T..... .C........ .......... .G....T.C.
HOBLERLILY  .......... .......... ....T..... .C........ .......... .G....T.C.

                                                                           180
HUMAN       GAGAGTCTGT GCCTCAGTGC CTCCTGCACT GAGTGGCATC CAAGTGGCGC CACCTCTCAG
PATTI       .......... .......... ........T. ...C...... .......... ..........
BROOK       .......... .......... ........T. ...C...... .......... ..........
ROGER       .......... .......... ........T. ...C...... .......... ..........
SISWI       .......... .......... ........T. ...C...... .......... ..........
ROSEMARY    .......... .......... ........T. ...C...... .......... ..........
1228        .......... .......... ........T. ...C...... .......... ..........
DOC         .......... .......... ........T. ...C...... .......... ..........
BELLA       .......... .......... ........T. ...C...... .......... ..........
APOLLO BOB  .......... .......... ........T. ...C...... .......... ..........
HERBIE      .......... .......... ........T. ...C...... .......... ..........
HOBLERLILY  .......... .......... ........T. ...C...... .......... ..........
```

```
                                                                      240
HUMAN       GTTCCTGGGT GGGCAAGAAG CGGTGCACGT CCAGGGCCTC CCACCAGGGC TGGCAGCCC-
PATTI       .......... .......... .A..A..... .......... A......--. .........C
BROOK       .......... .......... .A..A..... .......... A......--. .........C
ROGER       .......... .......... .A..A..... .......... A......--. .........C
SISWI       .......... .......... .A..A..... .......... A......... .........-.
ROSEMARY    .......... .......... .A..A..... .......... A......... .........-.
1228        .......... .......... .A..A..... .......... A......... .........C
DOC         .......... .......... .A..A..... .......... A......... .........-.
BELLA       .......... .......... .A..A..... .......... A......... .........-.
APOLLO BOB  .......... .......... .A..A..... .......... A......... .........-.
HERBIE      .......... .......... .A..A..... .......... A......... .........-.
HOBLERLILY  .......... .......... .A..A..... .......... A......--. .........C


                                                                      300
HUMAN       AGGTATGTGC AGTGCTTGGG CCTGCCCCGC CCCGTGACCC CTGA-CTCTG CCCCCAGAGC
PATTI       .......... G......... ....T..... .........T. .......... ..........
BROOK       .......... G......... ....T..... .........T. .......... .........--
ROGER       .......... G......... ....T..... .........T. .......... ..........
SISWI       .......... G......... G.CTGT.... .........T. ....G..... ..........
ROSEMARY    .......... G......... G.CTGT.... .........T. .......... ..........
1228        .......... G......... G.CTGT.... .........T. .......... ..........
DOC         .........T G......... G.CTGT.... .......... .......... ..........
BELLA       .........T G......... G.CTGT.... .......... .......... ..........
APOLLO BOB  .......... G......... G.CTGT.... .........T. .......... ..........
HERBIE      .......... G......... G.CTGT.... .........T. .......... ..........
HOBLERLILY  .......... G......... ....T..... .........T. .......... ..........


                                                                  346
HUMAN       CCTCGGCAGT CTCTCTTCTT CTACCCGTCC TACCCAGACG AGGTCC
PATTI       .......... .......... .......... .......... ....??
BROOK       ------.... .......... .......... .....????? ??????
ROGER       .......... .......??. T.???????? ?????????? ??????
SISWI       .......... .......... ...??????? ?????????? ??????
ROSEMARY    .......... .......... .......... .......... ......
1228        .......... .......... .......... .......... ......
DOC         .......... .......... .......... .......... ......
BELLA       .......... .......... .......... .......... ......
APOLLO BOB  .......... .......... .......... .......... ......
HERBIE      .......... .......... .......... .......... ......
HOBLERLILY  .......... .......... .......... .......... ......
```

Appendix 9: Lysozyme Intron 3 sequence alignment

```
PPYINT3          GTATGTTTTA AGTGTTAAAA GGGAAAACTA TTTTACTCTA CTGTTGATAC  50
Patti            .......... ......G... .......... .......... ..........
Roger            .......... .......... .......... .......... ..........
Doc              .......... .......... .......... .......... Y.........
Herbie           .......... .......G.. .......... .......... ..........
Mark             .......... ...A.A.... .......... .A........ Y.........
Bella            .......... .......G.. .......... .......... ..........

PPYINT3          ATACAACGAG AGCAGACTTT TAAAGACCAA AGTATGCTAA TGACACCTAA  100
Patti            ...T...... .......... .......... .......... ..........
Roger            .......... .......... .......... .......... ..........
Doc              .......... .......... .......... .......... ..........
Herbie           ....T..... .......... .......... .......... ..........
Mark             .......... ........CA .......... .......... ..........
Bella            .......... .......... .......... .......... ..........

PPYINT3          AAATTGCAGC TTTTGGCTTA TGCTAAATGA TGTATTACCT ACATCCTTGA  150
Patti            .......... .......... .......... .......... ..........
Roger            .......... .......... .......... .......... .........A
Doc              .......... .......... .......... .......... ........A.
Herbie           .......... .......... .......... .......... ..........
Mark             .......... .......... .......... .......... ......G...
Bella            .......... .......... .......... .......... ..........

PPYINT3          AGAGACAATC TACTTTAACT GATCCAGAAT CGTACTCTTT TACTCCTCAA  200
Patti            .......... .....G.... .......... .......... ..........
Roger            .......... .......... .......... ....G..... ..........
Doc              .......... .......... ......T... .......... .........T
Herbie           .......... ......G... .......... .......... ..........
Mark             .......... .......... .......... ...T...... ..........
Bella            .......... ......G... .......... .......... ..........

PPYINT3          TTTAATTTAG GGGAGGTCTA GAGTTTTAAG ATGCTTCACA CTCTATCAGT  250
Patti            .......... .......... .......... .......... ..........
Roger            .......... .......... .T........ .......... ..........
Doc              .........T T......... .......... .......... ..........
Herbie           .......... .......... .......... .......... ..........
Mark             .......... .......... .......... .......... ..........
Bella            .......... .......... .......... T......... ..........

PPYINT3          TCCTTGTCAT ATTATGAAAT TCCTTTTAGA ATAAGTAAAT GTGGTCCGGG  300
Patti            .......... .......... .......... .......... ..........
Roger            .......... .......... .......... .......... ..........
Doc              ........T. .......... .......... .........G ..........
Herbie           .......... .......... .......... .......... ..........
Mark             .......... .......... .......... .......... .....G....
Bella            .......... .......... .......... .......... ..........

PPYINT3          CACGGTGGCA CACGCCTGTA ATCCCAGCAC TCTGGGAGAC TGAGGCAGAT  350
Patti            .......... .......... .......... .......... ..........
Roger            .......... .......... .......... ...G...... ..........
Doc              .......... .......... .....T.... .......... ..........
Herbie           .......... .......... .......... .......... ..........
Mark             .......... .......... .......... .......... ..........
Bella            .......... .......... .......... .......... ..........
```

Appendix 9: Lysozyme Intron 3 sequence alignment (continued)

```
PPYINT3            GGATCACCTG AGGTCAGGAG TTCGAGACCA GCCTGCCTAG CATGGCAAAA  400
Patti              .
Roger              .......... .......... .......... .......... ..........
Doc                .......... .......... .......... .......... ..........
Herbie             .......... .......... .........C .......... ..........
Mark               .......... .......... .......... .......... ..........
Bella              .......... .......... .......... .......... ..........

PPYINT3            CCCCATCTCT ACTAAAAATA CAAAAAATTA GCTGGGTGTG GTGTCAGGTG  450
Patti
Roger              .......... .......... .......... .......... ..........
Doc                .......... .......... .......... .......G.. ..........
Herbie             ....
Mark               .......... .......... .......... .......... ..........
Bella              ........

PPYINT3            CCTGTAATCC CAGCTCGGGA GGCTGAGGCA GGAGACTTGC TTGAACCCGG  500
Patti
Roger              .......... .......... .......... .......... ..........
Doc                .....A.... .......... .......... .......... ..........
Herbie
Mark               .......... .......... .......... .......... ..........
Bella

PPYINT3            GAGGTGGAGG TTGCAGAGGA CTGCACCACT GCACTTCAGC CTGGGCAACA  550
Patti
Roger              .......... .......... .......... .......... ....GTC...
Doc                .......... .......... .......... .......... ..........
Herbie
Mark               ........C. .......... .......... .......... ........S.
Bella

PPYINT3            GAGTGAGACT CTGTCTCAGA TAAATAAATA AATAAATAAT GTGGAATTCA  600
Patti
Roger              .......... .......... .......... .......... ..........
Doc                .......... .......... .......... .......... ..........
Herbie
Mark               .......... .......... .......
Bella

PPYINT3            CTTTGCAGTT GCTGCTGTAC AACGCACATT ACTCAATCTT TATGTTCAGC  650
Patti
Roger              .........T ...TT..... A......T.A A..A...... .T...AT...
Doc                .......... .......... .......... .......... ........GT
Herbie
Mark
Bella

PPYINT3            ATTCTATGTT CTACTGAGAA ATTCGGGTAG GAGTGAAGTA TTTTGTATAC  700
Patti
Roger              ..A......A A...TT.A.. ..TT..TT.A ....T....A ...
Doc                C.......... .......... .......... .......... ..........
Herbie
Mark
Bella
```

Appendix 9: Lysozyme Intron 3 sequence alignment (continued)

```
PPYINT3            ATATCTTCAT TTAATAAATA GCAATAGCTG GGTCTATCTT ACTATTTTAT  750
Patti
Roger
Doc                .......... .......... .......... .......... ..........
Herbie
Mark
Bella
```

```
PPYINT3          CTATTGATAA AATATTGTGT CCCCCCAGGG AGTGWGGTGT GTGTATATCA  800
Patti
Roger
Doc              ........T. ..TT.....A ......T.AA ..A....... T...AT....
Herbie
Mark
Bella

PPYINT3          CAGAGGAGAT GTGTTTTCCC CTCACCCCAT GCGCCACCTC TTCTTCTGCA  850
Patti
Roger
Doc              .A......AA ...TT.A... .TT..TT.A. ...T....A. ..
Herbie
Mark
Bella

PPYINT3          G                                                      851
Patti
Roger
Doc
Herbie
Mark
Bella
```

Appendix 10: SRY sequence alignment

```
SRY             CTCTAGGGGG TAGACTGGTT GGGCGGGGTG AGAGGGGTGT TGGGGGCGGA  50
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             GAAATGAAAG TTTCATTACA AAAGTTAACG TAACAAAGAA TCTGGTAGTA  100
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             GTGAGTTTTG GATAGTAAAA TAAGTTTCGA ACTCTGGCAC CTTTCAATTT  150
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             TGTCTCATCC TCCTTGTTTT TCACAATGCA ATCATATGCT TCTGCTATGT  200
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             TAAGCGTATT CAACAGTGAT GATTACAGTC CAGCTGTGCA ACAGAATATT  250
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             CCCGCTCTCC GGAGAAGCTC TTCCTTCATT TGCACTGAAA GCTATAACTC  300
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             TAAGTATCAG TGTGAAACGG GAGAAAACAG TAAAGGCAGC GTCCAGGATA  350
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             GAGTGAAGCG ACCCATGAAC GCATTCATCG TGTGGTCTCG CGATCAGAGG  400
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             CGCAAGATGG CTCTAGAGAA TCCCAAAATG CGAAACTCAG AGATCAGCAA  450
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........
```

Appendix 10: SRY sequence alignment

```
SRY             GCAGCTGGGA TACCAGTGGA AAATGCTTAC TGAAGCCGAA AAATGGCCAT  500
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             TCTTCCAGGA GGCACAGAAA CTACAGGCCA TGCATAGAGA GAAATACCCG  550
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             AATTATAAGT ATCGACCTCG TCGGAAGGCG AAGATGCTGC AGAAGAGTTG  600
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             CAGTTCGCTT CCCGCAGATC CCGCTTCGGT ACTCTGCAGC GAAGTGCTGC  650
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             AACTGGACAA CAGGTTGTAC AGGGATGACT GTACGAAAGC CACACACTCA  700
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             AGACTGGAGC ACCAGCTAGG CCACTTACCG CCCATCAACA CAGCCAGCTC  750
Mark            .......... .......... .......... .......... ..........
Herbie          .......... .......... .......... .......... ..........
Ranto           .......... .......... .......... .......... ..........
Stan            .......... .......... .......... .......... ..........
ApolloBob       .......... .......... .......... .......... ..........

SRY             ACCGCAGCAA CGGGACCGCT ACAGCCACTC GACAGAGCTG TAC         793
Mark            .......... .......... .......... .......... ...
Herbie          .......... .......... .......... .......... ...
Ranto           .......... .......... .......... .......... ...
Stan            .......... .......... .......... .......... ...
ApolloBob       .......... .......... .......... .......... ...
```

Appendix 11. COII and spacer sequence alignment

```
GBCOIILY      GTATATTATG GCCAATGCTC AGAAATCTGT GGAGCTAACC ACAGCTTTAT GCCTA
SISWI         .......... .......T.. ...G...... .......... .......... .....
RUBY          .......... .......... .......... .......... .......... .....
ABIGAIL       ..G..C.... .......... .......... .....C.... .......... .....
SUMCOII       ..G..C.... .......... .......... .......... .......... .....
ROSEMARY      .G........ .......... .......... .......... .......... .....
PATTI         .G........ ....G..T.. ...G...... .......... .T....C... .....
MARK          .......... .......T.. ...G...... .......... ......A... .....
HOBLERLILY    .......... .......T.. ...G...... .......... .T........ .....
BROOK         .......... .......T.. ...G...... .......... .T....A... .....
RANTO         .......... .......... ...G...... .......... .T....A... .....
"A"           .......... .......... .......... .......... .T....A... .....
STAN          .......... .......... ...G...... .......... .T....A... .....

GBCOIILY      TCGTC CTAGAACTAA TCCCCCTAAA AATCTTCGAA ATAGGACCCG TATTCGCTTT
SISWI         ....T .......... .T........ .......... ........T. .....A....
RUBY          ..... .......... .......... .......... .......... .....A....
ABIGAIL       ..... ......T... .......... .......... .....G.... .....A....
SUMCOII       ..... .......... .......... .......... .....G.... .....A....
ROSEMARY      ...G. .......... .......... .......... .......... .....A....
PATTI         ...GT .......... .T........ .......... ........T. ..........
MARK          ....T .......... .T........ .......... ........T. .....A....
HOBLERLILY    ....T .......... .T........ .......... ........T. .....A....
BROCK         ....T .......... .T........ .......... ........T. ..........
RANTO         ....T .......... .T........ .......... ........T. ..........
"A"           ....T .......... .T........ .......... ........T. ..........
STAN          ....T .......... .T........ .......... ........T. ..........

GBCOIILY      ATAACTCTTC CACCCCCCCC CCCCATCCTA CCTCCTTTCC TGAGGC
SISWI         ...G
RUBY          .......... .......... ATCCTACCTA CTTTCCTGAG GCC
ABIGAIL       .....C..C. .......A.A A.CCATCCTA CCCC
SUMCOII       ......TCC. .CA.....A. AACCCATCCT ACCCCCTTTC CTGAGGCC
ROSEMARY      .......... ..A....... ..CCATCCTA CCTCCTTTCC TGAGGTC
PATTI         ...G
MARK          ...G
HOBLERLILY    ...G
BROOK         ...G
RANTO         ...G
"A"           ...G
STAN          ...G
```

References

Adams, J.M. and F.I. Woodward. 1992. The past as a key to the future; the use of palaeoenvironmental understanding to predict the effects of man on the biosphere. Advances in Ecological Research, 22: 257-314

Adams, J.M. 1993. Towards a Better Vegetation System for Global Mapping and Monitoring. Background paper for the UNEP-HEM Charlottesville Workshop, January 1993.

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

Andrews, P., Cronin, J.E., 1982 The relationship of *Sivapithecus* and *Ramapithicus* and the evolution of the orang-utan., Nature, 297: 541-545

Cabot, E.L., A.T. Beckenbach, 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Computational and Applied Bioscience. 5:233-234.

Cann, R.L., A.C. Wilson, 1983, Length Mutations in Human Mitochondrial DNA. Genetics 104: 699-711

Cann, R.L., W.M. Brown, A.C. Wilson, 1984, Polymorphic Sites and the Mechanism of Evolution in Human Mitochondrial DNA. Genetics, 106:479-499

Chesner, C.A., W.I. Rose, A. Deino, R. Drake, J.A. Westgate, 1991, Eruptive history of Earth's Largest Quaternary caldera (Toba Indonesia) clarified. Geology 19:200-203

Clayton, D.A., 1982, Replication of Animal Mitochondrial DNA. Cell 28:693-705

Collura R.V. and Stewart C., 1995 Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. Nature 378(6556):485-489

Crandall, K.A., 1994, Intraspecific Cladogram Estimation: Accuracy at Higher Levels of Divergence. Systamatic Biology 43(2) 222-235.

Crandall, K.A., A.R. Templeton, 1993, Empirical Tests of Some Predictions From Coalescent Theory with Applications to Intraspecific Phylogeny Reconstruction. Genetics 134: 959-969.

deBoer L.E.M., 1982 Genetics and conservation of the orangutan. The orangutan. Its biology and conservation. (deBoer ed.) Dr. W. Junk Publishers, The Hague

Dobzhansky, Th., 1937, Genetic nature of species differences, Amer. Natur.

71:404–420.

Drawhorn, G.M., 1994 The systematics and paleodemography of fossil orangutans (Genus Pongo). Ph.D. thesis University California Davis.

Galdikas, B.M.F. 1981 Modern adaptations in Orangutans? Nature 291: 266

Galdikas, B.M.F., 1982 Orang utans as seed dispersers at Tanjung Puting, Central Kalimantan: implications for conservation. The orang utan. Its biology and conservation. pp 285 - 298(L.E.M. de Boor ed) Dr. W. Junk Publishers, The Hague

Galdikas, B.M.F., 1984 Adult female sociality among wild orangutans at Tanjung Puting Reserve. Female primates: Studies by women primatologists, pp217 - 235, Alan R. Liss, Inc. New York, NY

Galdikas B.M.F., 1985, Sub-adult male orangutan sociality and reproductive behavior at Tanjung Puting. American Journal of Primatology. 8:87-99

Galdikas B.M.F., 1985, Adult Male Sociality and Reproductive Tactics Among Orangutans at Tanjung Puting. Folia Primatologica, 45: 9-24

Galdikas, B.M.F., 1988, Ornagutan diet, range, and activity at Tanjung Puting, Central Borneo. International Journal of Primatology, 9:1:1-35

Galdikas, B.M.F., Wood, J.W.,1990, Birth spacing patterns in humans and apes. American Journal of Physical Anthropology. 83:185-191

Garner, K.J., O.A., Ryder, 1992, Some Applications of PCR to Studies in Wildlufe Genetics. Symposium of the Zoological Society of London 64:167-181.

Garza J.C., Freimer N.B., 1996, Homoplasy for Size at Microsatellite Loci in Humans and Chimpanzees. Genome Research 6:211-217

Goldberg, T.L., M. Ruvolo, 1997, The Geographic Apportionment of Mitochondrial Genetic Diversity in East African Chimpanzees, *Pan troglodyte schweinfurthii.* Molecular Biology and Evolution. 14 (9):976-984

Hartl, D.L., and Clark, A.G., 1989, *Principles of Population Genetics*, 2nd Ed., Sinauer Associates, Sunderland, Massachusetts.

Horai, S., Satta, Y., Hayasaka, K., Kondo, R., Inoue, T., Ishida, T., Hayashi, S. and Takahata, N., 1992, Man's place in Hominoidea revealed by mitochondrial DNA genealogy, *J. Molec. Evol.* 35:32-43.

Janczewski, D.N., Goldman, D., and O'Brien, S.J., 1990, Molecular genetic divergence of orang utan (*Pongo pygmaeus*) subspecies based on isozyme and two–dimensional gel

electrophoresis, *J. Heredity* 81:375–387.

Jolly C.J., Oates J.F., Disotell T.R., 1995, Chimpanzee Kinship. Science 268: 185-186

Linnaeus, C., 1735, *Systema Natuae sive Regna Tria Naturae Systematice Proposita per Classes, Ordines, Genera, and Species.* (Translated reprint 1964 Engle-Ledeboer and Engle Nieuwkoop B. De Graaf

Locke, J., 1690 (reprinted in English 1956), Essay concerning human understanding. Gateway Editions, Henry Regenry Company, Chicago

Mallet, J., 1995, A definition for the modern synthesis. Trends in Ecological Evolution. 10:(7) 294-299

Martin, L., Andrews, P.,1993, Renaissance of Europe's ape. Nature. 365: 494
Mayr, E., 1940, Speciation phenomena in birds, *Amer. Natur.* 74:249–278.

Mayr E., 1965, Animal Species and Evolution. The Belknap Press of Harvard University Press, Cambridge, Massachusetts

Miles, L., 1995, Chantek. The Neglected Ape. (Nadler, et al. ed's) Plenum Press

Morin, P.A., Wallis, J., Moore, J.J., Chakraborty, R., and Woodruff, D.S., 1993, Non-invasive sampling and DNA amplification for paternity exclusion, community structure, and phylogeography in wild chimpanzees, *Primates* 34:347-356.

Moya Sola, S., Kohler, M., 1993, Recent discoveries of *Dryopithecus* shed new light on evolution of great apes. Nature. 365: 543-545

Muir C., Galdikas B.M.F., Beckenbach A.T., 1995, Genetic Variability in Orangutans. The Neglected Ape (Nadler R., et al. ed.s) Plenum Press, New York, 1995

Muir,C.Cam, Galdikas, Birute M.F., Beckenbach, Andrew T., 1998a Is there sufficient evidence to elevate the orangutan of Borneo and Sumatra to separate species? J Molecular Evolution 46:378-381

Muir, C. Cam, Watkins, Russel F., Ha, The 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 (accepted)

Muir, C. Cam, Ha, The T., Watkins, Russel F.,Vankoeveringe, Michael, 1998c Applying bathometric data to population genetics of the orangutan through 3-D map simulations in Geoenironmental mapping: Methods and Applications (accepted)

Muir, C. Cam, Galdikas, Birute.M.F., Beckenbach, Andrew.T., Inferences into population structure and migratory history of the orangutan of Borneo and Sumatra from ND3 sequence data. Evolution (in review)

Mullis, T., and Faloona, F.A., 1987, Specific synthesis of DNA *in vitro* via a polymerase–catalyzed chain reaction, *Methods Enzymol.* 155:335–350.

Peltier, R.W., 1994, Ice age paleotopography. Science 265: 195 - 201

Rampino M.R., S. Self, 1992, Volcanic winter and accelerated glaciation following the Toba super-eruption. Nature, 359:50-52

Ruvolo M., Pan D., Zehr S., Goldberg T., Disotell T., von Dornum M., 1994, Gene trees and hominoid phylogeny. Proc. Natl. Acad. Sci. USA 91:8900-8904

Ryder, O.A., and Chemnick, L.G., 1993, Chromosomal and mitochondrial DNA variation in orang utans, *J. Heredity* 84:405–409.

Seuanez, H.N., H.J. Evans, D.E. Martin, and J. Fletcher, 1979, An inversion in Chromosome 2 that distinguishes between Bornean and Sumatran orangutans. Cytogenetics and Cell Genetics. 23:137-140

Smith, R.J., Pilbeam, D.R., 1980, Evolution of the orangutan., Nature, 284:447-448

Steinberg, S., Cedergren, R., 1995, A correlation between N^2 - demethylguanosine presence and alternative tRNA conformers. RNA 1:886-891

Tautz D and Renz M, 1983, An optimized freeze-squeeze method for the recovery of DNA fragments from agarose gels. Analytical Biochem 132:14-19.

Temerin, L.A., 1980, Evolution of the orangutan., Nature, 288:301.

Uchida A., 1996, What We Don't Know About Great Ape Variation. TREE 11 (4) 163-168

Van der Kaars. Sander, 1990, Late Quaternary Vegatation nd Climate of Australia as Reflected by the Palynology of Eastern Indonesian Deepsea Piston-cores. Ph.D. Thesis University of Amsterdam

Van der Kaars, S., Dam, R., 1996, Vegetation and Climate change in West-Java, Indonesia during the last 135,000 years. Quaternary International 37: 67-71

Vigilant, L., Pennington, R., Harpending, H., Kocher, T.D., and Wilson, A.C., 1989, Mitochondrial DNA sequences in single hairs from a southern African population, *Science* 253:1503-1507.

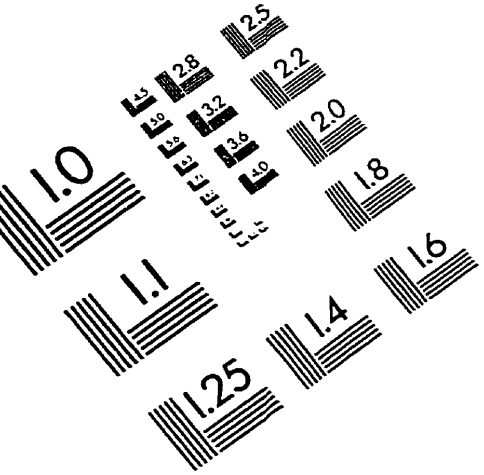
Walsh, P.S., D.A. Metzger and R. Higuchi, 1991 Chelex 100 as a medium for PCR based typing from forensic material. Biotechniques 10:506-513

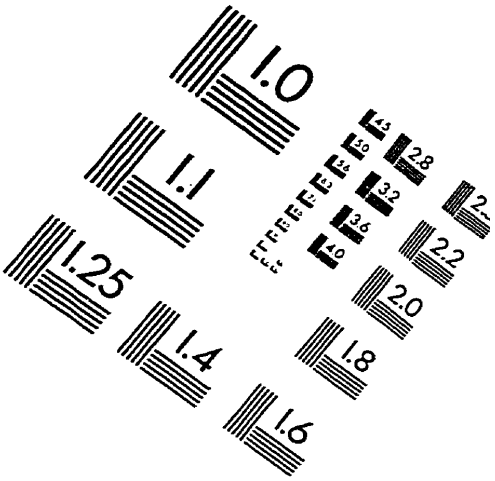
Whitfield, S.L., R. Lovell-Badge, P.N. Goodfellow, 1993 Rapid sequence evolution of the mamalian sex-determining gene SRY. Nature 364:713-717

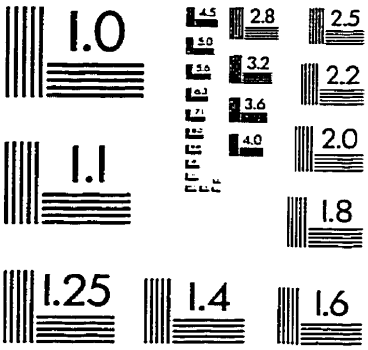
Xu X., Arnason U., 1996, The Mitochondrial DNA Molecule of Sumatran and a Molecular Proposal for Two (Bornean and Sumatran) Species of Orangutan. J. Mol. Evol. 43: 431-437

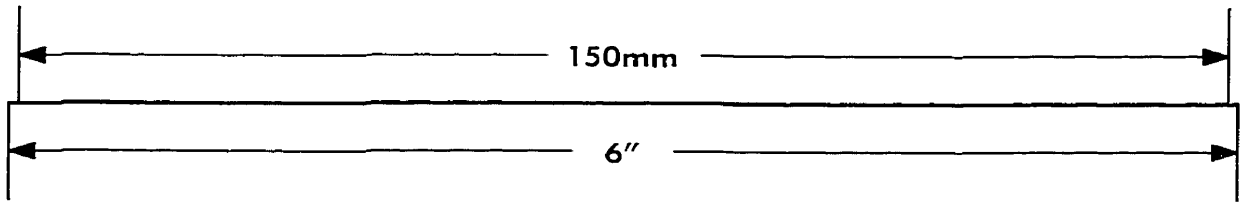
Zhi L., Karesh W.B., Janczewski D. N., Frazier-Taylor H., Sajuthi D., Gombek F., Andau M., Martenson J.S., O'Brien S.J., 1996, Genomic Differentiation Among Natural Population of Orang-utan (*Pongo pygmaeus*). Curr. Biol. 6 (10) 1326-1336

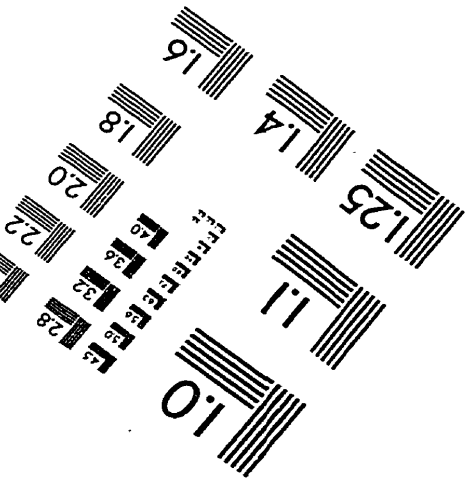
IMAGE EVALUATION TEST TARGET (QA-3)

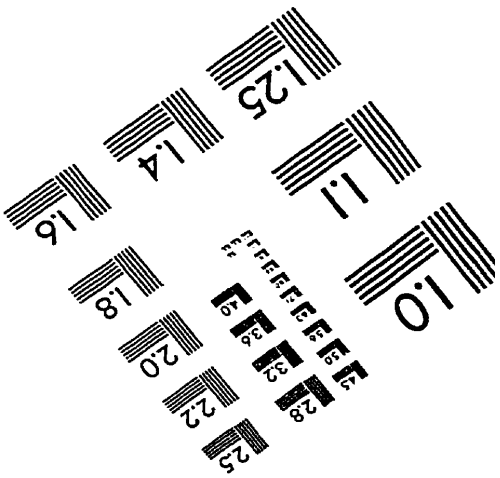












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989