EXPLORING HIDDEN GENETIC DIVERGENCE WITHIN SUNDA COLUGOS BY MEANS OF NOVEL DNA CAPTURE METHODS AND NEXT GENERATION SEQUENCING

A Senior Scholars Thesis

by

VICTOR C. MASON

Submitted to the Office of Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

April 2010

Major: Biomedical Science

EXPLORING HIDDEN GENETIC DIVERGENCE WITHIN SUNDA COLUGOS BY MEANS OF NOVEL DNA CAPTURE METHODS AND NEXT GENERATION SEQUENCING

A Senior Scholars Thesis

by

VICTOR C. MASON

Submitted to the Office of Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

Approved by:

Research Advisor: William J. Murphy
Associate Dean for Undergraduate Research: Robert C. Webb

April 2010

Major: Biomedical Science

ABSTRACT

Exploring Hidden Genetic Divergence Within Sunda Colugo by Means of Novel DNA Capture Methods and Next Generation Sequencing. (April 2010)

Victor C. Mason
Department of Biomedical Science
Texas A&M University

Research Advisor: Dr. William J. Murphy
Department of Veterinary Integrative Biosciences

It has been the goal of biologists to catalog and protect genetic diversity and variation among biological organisms. The amount of diversity cataloged is growing every year. In the twelve years between the latest two publications of *Mammal Species of the World*, the number of mammalian species increased from 4998 to 5339 (~7%). This number is expected to increase substantially, especially with the advent and application of new genomic approaches to assess levels of species diversity. This increased diversity is partially due to increased taxonomic investigation in Southeast Asia, which is known for being a hot spot of species richness. This richness has been shown in recent years to be continually threatened by human induced habitat loss, as is the case of a poorly known group of mammals, the flying lemurs, or colugos. The colugo is a small arboreal mammal that inhabits more than fifty islands in the SE Asian archipelago and adjacent mainland areas of the Malay Peninsula, Thailand and Vietnam. The colugo has extremely inefficient terrestrial locomotor capabilities, which isolate the colugo to forested areas, where it is capable of gliding over one hundred meters between trees.

This study proposes a molecular phylogenetic analysis of the Sunda colugo (*Galeopterus variegatus*) to redefine the evolutionary relationships between disjunct populations of this poorly understood mammal, using a novel DNA capture method to isolate degraded mtDNA fragments from museum samples, by hybridization to DNA fragments derived from a modern colugo genome. The results demonstrate extremely efficient cross-species capture of mtDNA sequences as great as 10-15% divergent from the probe, combined with Next Generation Sequencing Technologies to obtain high depth of coverage of hybridized sequences. Phylogenetic results indicate the widespread presence of species-level taxonomic units both within and between the islands of the Southeast Asian archipelago. This novel approach to ancient DNA capture has potentially broad implications for the conservation of this enigmatic mammal, and further suggest that vicariant evolutionary analysis of colugos will be invaluable for defining the biogeographic history of the SE Asian archipelago.

ACKNOWLEDGMENTS

I would like to thank the National Science Foundation for the opportunity and funding for this project. I would like to give a special thank you to Dr. William Murphy for his extensive time, help, and guidance throughout this whole process, with an extended thanks to Gang Li, Brian Davis, Jan Janecka, and Colleen Fisher for their help in the lab. Thanks to Kristofer Helgen and the Smithsonian Institution for providing the museum samples for this study.

NOMENCLATURE

IUCN	International Union For Conservation Of Nature
kb	Kilobases
mtDNA	Mitochondrial DNA
mya	Million Years Ago
NGST	Next Generation Sequencing Technologies
PCR	Polymerase Chain Reaction
RT	Room Temperature
Sunda	Sundaland
~	About Equal To

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Colugos are mammals classified in the Order Dermoptera, which in Greek translates to "skin-wing", which is fitting for the mammal with the largest patagium (gliding membrane). There are currently only two extant colugo species classified in the order: the Sunda Colugo, *Galeopterus variegatus*, and the Philippine Colugo, *Cynocephalus volans,* both classified in the Family Cynocephalidae (Lim 2007). Though colugos are also referred to as flying lemurs, they actually cannot fly (like bats); they glide, and they are not true lemurs (which are primates). This underscores the confusion surrounding this group of mammals that are the closest living relative of primates (Janecka et al. 2007). Colugos are extremely proficient gliders; one has even been recorded gliding up to one hundred and thirty six meters, a truly magnificent feat (Lim 2007). Despite their prowess as gliders, their magnificence fades when on the ground; colugos are extremely clumsy, tending to flop around rather than move with any sort of meaningful progression. This is aided by the fact that their hind legs cannot even hold up their own body weight (Lim 2007). This hindrance has imprisoned this arboreal mammal to the ever-dwindling, connected stretches of forest, preventing them from traversing any type of open grassland or savannah biome.

This thesis follows the style of *Genome Research*.

Colugos are distributed throughout Southeast Asia and spread across many islands, including Borneo, Sumatra, Java and the Philippines, which make up the geographical region known as Sundaland. This situation happens to be very similar to that of the Galapagos Islands where Darwin managed to identify 13 different species of finch. Genetic and reproductive isolation may even be more profound between the different colugo populations because something as simple as a river, or a human induced gap in the forest, creates an impenetrable barrier for these arboreal mammals (Lim 2007). If these simple and common barriers are impassable, it is easy to presume that multiple species can be present on the same mountainous island. Even neglecting the potential variation within one island, the broad distribution of colugos throughout Southeast Asia, coupled with their limited ability to disperse, suggests there are many more cryptic species of colugo spread throughout the SE Asian mainland and islands systems. These hidden species could be formed from separate founder events coupled with the effects of genetic drift in forest refugia throughout Sundaland. A founder event occurs when a fraction of a population is isolated from the remaining ancestral population, and is often representative of a small fraction of the larger populations' gene pool. Since this population is smaller (of finite size) and isolated, genetic drift has a larger impact on the frequencies of all alleles at all loci for each successive generation. These repeated founder events and population subdivisions, whether geological or man-induced, potentiates increased evolutionary divergence between small, isolated populations (Templeton 2006).

There has been very limited genetic or ecological research performed on colugos. Recent genetic research based on a limited number of isolated populations (Java, Borneo, Singapore and southern Thailand) suggests deep genetic divergences that are equivalent to species level differences (Janecka et al. 2008). Within *Galeopterus variegatus* only four subspecies are currently recognized by modern taxonomy (Stafford and Szalay 2006). This likely underestimates the potential genetic divergence considering the Sunda colugo is distributed throughout more than 50 islands, where each could provide a vicariant mechanism for speciation.

Dwarf populations of colugos are present on various islands and mainland areas. Dwarf populations of the Sunda colugo have been cataloged on nine isolated populations on islands south, east, and southwest of the Sumatran peninsula (Ruggeri and Etterson 1998). These dwarf populations present interesting phylogenetic questions. These species could be representative of a single historical species that has been isolated to their respective locales due to rising sea levels and forcing them to retreat into a more limited domain maintaining similar genetic divergence. Alternatively, these dwarf forms might also be the outcome of multiple forms of convergent evolution, where dwarfism may evolve at a very rapid rate due to selection on one or a few genes. The relationships of these different dwarf populations to one another, and to the Sunda colugo population as whole, is of interest to understanding the evolutionary plasticity of this trait. Two of these dwarf populations are present in our current sample set: one from the Natuna Islands, east of Sumatra, and another from Langkawi Island, west of Sumatra.

Currently the International Union for Conservation of Nature (IUCN) classifies the Sunda colugo as a taxon of "least concern," even when the impact of human induced habitat destruction on these mammals has potentially already been devastating. Making this judgment requires neglecting the multiple examples of genetic variation and embracing the fact that the world is rather ignorant of this group of mammals. Without a thorough study of the actual variation between different populations and regions we cannot pursue any sort of meaningful attempt for conservation. This is why coming to a better understanding of these mammals is so pressing. If there are multiple species of the colugo, this study might dramatically increase conservation efforts for this "increased" biodiversity, potentially classifying some species as endangered, and preventing further loss of millions of years of divergence.

Vicariance/island biogeography

With their broad distribution across multiple islands and mainland, colugos would seem to be a model organism for vicariant analysis. Vicariant speciation events occur when a population of organisms is divided due to a physical barrier, and this separation results in a speciation event. A trademark example that produces vicariant speciation is rising sea levels (Lomolino et al. 2006). During the Late Pleistocene, sea levels in SE Asia fluctuated by over 100 meters; falling to such an extent that someone could travel by foot from Malaysia to East Java (Harrison et al. 2006; Woodruff 2003). SE Asia is well known for its species richness throughout the archipelago as well as having a complex geographic history (Schipper et al. 2008). These reasons make mapping the sea levels

through this region's history necessary to understanding these organisms' potential dispersal options. If there was any chance for dispersal, then this must be taken into account so that historical phylogenetic patterns can be accurately interpreted.

Ancient DNA

Analysis of ancient DNA (i.e. from museum specimens) has become increasingly valuable for the analysis of extinct and contemporary forms of a species, or species that are rare and difficult to catch in the wild (Millar et al. 2008). One approach that makes this a feasible and common practice is a technique involving utilization of next generation sequencing technologies and a rather unique hybridization-based selection procedure. Next generation sequencing technologies (NGSTs) allow for the random replication and sequencing of every piece of DNA in a sample, in some cases even producing several gigabases of data (Millar et al. 2008). These billions of base pairs sequenced allow for never before seen depth of sequencing, which provides the interpreter with the ability to rule out incorrect sequences and provides great depth of coverage across genomes, even where the amount of remaining DNA in a sample is relatively small in proportion to bacterial or other contaminating DNA. This depth of coverage is so important because in the case of mtDNA, which this study is largely based on, it provides a means to confidently rule out numts (segments of mitochondrial DNA transposed into the nuclear genome) that commonly plague mtDNA analyses. Further benefits of the depth of sequencing involve corrections of cytosine mutations which are prevalent in ancient DNA samples (Millar et al. 2008).

This thesis develops a comprehensive method that will capture both ancient mtDNA (and potentially nuclear DNA) from several museum samples of the Sunda colugo that range from 60 to 150 years old. Coupling this with next generation sequencing technologies allows for venture into a largely unexplored area of ancient DNA retrieval. This procedure provides guidelines from the initial DNA extractions to the retrieval of selected products. This method not only has grand implications for the taxonomy and conservation of the Sunda colugo but also can be widely implemented for the historical analysis of both extinct and extant taxa. Immense historical tissue samples are stored in museums around the world, and the method presented here enables new opportunities for the non-invasive utilization of these samples allowing for expanded historical analyses throughout the field of genetics.

CHAPTER II

METHODS

This study proposes a complete and comprehensive method for novel mitochondrial DNA capture from ancient museum samples, though in principle it is also applicable for targeted recovery of nuclear DNA fragments. This procedure will be outlined in its entirety, from DNA extraction to next generation sequencing, and divided into five major sections: preparation of probe, ancient DNA sample preparation, DNA hybridization, DNA selection, and options for next generation sequencing technologies. This division of the procedure will allow for better understanding, and opportunities to state key checkpoints for confirmation of progress.

Preparation of probe

Mitochondrial enrichment

1.5 mg of tissue (in this case, liver preserved in ethanol) was used for mitochondrial enrichment and extraction of DNA. The tissue was homogenized in 1mL of prechilled homogenization buffer (30mM Tris-HCl, 1mM EDTA, 2.5 mM $CaCl_2$, 0.25 M sucrose) in a 1.5 mL tube using a pestle. The homogenate was spun at 3500 rpm's for 15 minutes in a 4˚C microcentrifuge to pellet the nuclear debris. The supernatant was removed and transferred to another tube. The nuclear pellet was resuspended in a second, 600 µl aliquot of cold homogenization buffer and spun at 4000 rpm's for 10 minutes. The supernatant was removed and added to the tube containing the first supernatant. The

tube with the combined supernatants was spun at max speed in a microcentrifuge (~13400 rpm's) for 30 minutes at 4˚C to pellet the mitochondria. The nuclear pellet was frozen at -20˚C to -80˚C for future use.

The mitochondrial pellet was resuspended in 200 µl of cell lysis solution (Puregene® D-5002, Gentram Qiagen No#158908) and 5 µl of Qiagen Protienase K solution (ABI No#4333793.) After vortexing gently, the solution was incubated overnight at 56˚C. Following incubation the tube was placed on ice for 7 minutes. Sixty seven microliters of protein precipitation solution (puregene D-5003) was added, vortexed for 20 seconds, and spun for 5 minutes at 8000 rpm. The supernatant was transferred into a fresh tube for the mitochondrial-enriched DNA to be precipitated. Two hundred microliters of 100% isopropanol was added, inverted 50 times, and spun for 15 minutes at 12000 rpms. The DNA pellet was washed with 200 µl of 70% ethanol and spun for 2 minutes at 12000 rpms. The ethanol was discarded and the pellet allowed to air dry for 5-10 minutes. 20 µl of TLE was added to resuspend the DNA pellet. The DNA sample was quantified using a Nanodrop machine and analyzed on a 1% agarose gel to confirm overall DNA integrity. The DNA was stored at -20˚C.

Amplification of mitochondrial enriched DNA

19 primer pairs were designed from the Sunda colugo reference mtDNA genome in GenBank (Acc. No. AJ128849.1) to generate a series of overlapping amplicons, via polymerase chain reaction (PCR) using mitochondria-enriched DNA, that would serve as

a capture hybridization probe. Each mitochondrial amplicon was amplified in five replicate 25 µl PCR reactions with 2 µl of mitochondrial enriched DNA template. For each PCR reaction a unique primer pair was used (0.25 µl of forward primer, 0.25 µl of reverse primer, from 20 µM stocks), Invitrogen Platinum Taq (0.10µl), platinum taq 10X buffer without $MgCl_2$ (2.5 µl), 50mM $MgCl_2$ (0.75 µl), and a 2.00µl of 10mM dNTP mixture. The PCR profile used a 2 minute hot start at 94˚C, and 40 cycles of denaturing for 15 seconds at 94˚C, annealing for 30 seconds at 60-50˚C, and extension for 1 minute at 72˚C. During the first 10 cycles the annealing temperature was decreased from 60˚C to 50˚C by 2˚C increments every 2 cycles. A 2 minute final extension at 72˚C completed the reaction. These were purified using Montage filters (Millipore, Cat# UFC7PCR50) to remove residual reaction reagents. These products were resolved on a 1% agarose gel for visualization of product and quantified using a Nanodrop reader. The 19 different products were pooled together evenly to a total of 1 µg of DNA based on their concentrations and product size.

Biotin labeling

One microgram of pooled mtDNA fragments was labeled with biotin, which allows for subsequent capture on strepavidin beads. The 1 µg aliquot was dried down to a volume of ~10 µl, and then brought to a volume of 16 µl with water. The DNA was denatured at 99˚C for 5-10 minutes, and placed immediately in an ice/water bath for ~20-25 seconds. 4 µl of Biotin-High Prime (Roche) was added to the denatured DNA during these 20 seconds, and incubated overnight at 37˚C. The reaction was stopped by adding 2 µl of

0.2M EDTA (pH 8.0) and incubation at 65˚C for 10 minutes. The labeled products are henceforth referred to as the capture probe. The probe was stored at -20˚C in a light-proof box, as the biotin is light sensitive.

Ancient DNA sample preparation

Museum DNA extraction

DNA was extracted from museum samples derived from non-invasively collected skull and tissue scrapings from cranial material of colugo specimens obtained from the Smithsonian Institution's National Museum of Natural History (NMNH). The twelve specimens were collected from throughout the SE Asian archipelago and varied from 50 – 150 years old (Table 1). 200 µl of cell lysis solution (Puregene® D-5002, Gentram Qiagen No#158908) and 5 µl of Qiagen Proteinase K solution (ABI No#4333793) were added to each of the samples. After gentle vortexing, the samples were incubated overnight at 56˚C. Following incubation the tubes were placed on ice for 7 minutes. 67 µl of protein precipitation solution (Puregene D-5003) was added, vortexed for 20 seconds, and spun for 5 minutes at 8000 rpm. The supernatants were then transferred to a fresh tube to precipitate the mitochondrial enriched DNA. 200 µl of 100% isopropanol was added, inverted 50 times, and spun for 15 minutes at 12000 rpms. The DNA pellets were washed with 200 µl of 70% ethanol and spun for 2 minutes at 12000 rpms. The ethanol was removed and the pellets were allowed to air dry for 5-10 minutes. 20 µl of TLE was added to resuspend the pellet. After resuspension the DNA was quantified

using a Nanodrop device and analyzed on a gel to check integrity. The products were stored at –20˚C

Table 1. Collection data and distribution of museum samples used in this study.

Index #	USNM ID#	Date Collected	Geographical Location
1	154600	25 May 1909	Mt. Salak, West Java
2	155363	6 Apr 1909	East Java
3	307553	28 Sep 1957	Mt. Brinchong, Malaysia
4	311297	17 Jul 1958	Langkawi Island, Malaysia
5	197203	2 Jul 1913	Labuan Klambu, Borneo
6	317119	23 Sep 1960	Sabah, Borneo
7	356666	8 Feb 1963	Amphoe Kapoe, S. Thailand
8	198051	12 Jan 1914	Kari Orang, Borneo
9	104600	7 Jul 1900	Natuna Islands
10	115605	20 Aug 1902	Rhio Archipelago, Sumatra
11	121749	12 Feb 1903	Batu. Islands, Sumatra
12	143327	12 Mar 1906	Pulo Rupat, NE Sumatra

Blunt-ending of museum DNA extracts

The museum DNA extracts were blunt-ended to enable efficient ligation of an oligonucleotide adapter to ligate to the ends of each sequence. The adapters were formed by self-ligation of two oligos: (ORM-28 and ORM-29; Peterson 1998). A 1X 20 µl master mix was made on ice, composed of 9.5µl of H_2O, 8µl of 5X reaction buffer, 2.0 µl dNTPs (100µM each), and 0.5 µl T4 DNA Polymerase (0.1 U μl^{-1}). Twenty microliters of the master mix was added to 20 µl of each sample DNA to make a total reaction volume of 40 µl. The mixture was incubated at room temperature for 10 minutes before stopping the reaction by heating to 70˚C for 10 minutes. Following incubation, the samples were purified by use of a Sephadex column (CS-200, Princeton

Seperations). This reaction was stored on ice and immediately followed by the ligation step.

Adapter ligation of mtDNA

A 150 µl reaction mix was prepared on ice using 15 µl 10X Ligase Buffer + ATP, (New England BioLabs [NEB] Cat# B0202S), 5.6µl ORM-29 (4.25 µg), 6.7 µl ORM-28 (5 µg) 1.0 µl DNA Ligase (NEB, Cat# M0202S), 32.0 µl museum DNA, and 89.7 µl H_2O. The 10x buffer was warmed at 37˚C prior to addition to the reaction mix to dissolve any precipitated ATP. The ligation was carried out at 16˚C for 19 hours in a water bath.

Amplification of adapter-ligated mtDNA

Five 50µl PCR reactions were set up with 5µl of blunt-ended, adapter-ligated museum DNA as a template. The 1x PCR reaction contained ORM 28 primer (0.18 µg), Invitrogen Platinum Taq DNA polymerase (0.50 µl), Platinum Taq 10X buffer without $MgCl_2$ (5 µl), 50mM $MgCl_2$ (1.50 µl), and a 10mM dNTP mixture (4 µl.) The PCR profile used a 1 minute hot start at 95˚C, and 30 cycles of denaturing for 15 seconds at 94˚C, annealing for 20 seconds at 58˚C, and extension for 1 minute at 72˚C. A 5 minute final extension at 72˚C completed the reaction.

Hybridization

The hybridization step involves annealing of the most complementary segments of ancient DNA samples to the probe. 2x Hybridization Buffer (1.5mM NaCl, 40mM

Sodium Phosphate solution [18.25 ml of 1M NaH_2PO_4 (monobasic), 77.75 ml of 1M Na_2HPO_4 (dibasic), and 4 ml of H_2O], 10mM EDTA, 10x Denhardt's solution, and 0.2% SDS) was made fresh before each hybridization procedure. A volume of 2x hybridization solution that is equal to the sum of the volumes of 100 ng of biotin labeled probe and 1ug of ancient museum DNA was combined together, not exceeding 15 µl. After gently mixing the tube, this 15 µl sample was overlaid with 50 µl of mineral oil. The DNA was denatured for 5 minutes at 99°C, and immediately transferred to a 65°C water bath and incubated for 50 hours. Throughout the incubation time period the temperature was lowered by ~2°C every 24 hours to a final temperature of 60°C. This "touch-down" approach was used to enhance retrieval of more divergent DNA sequences relative to the capture probe.

Selection

Primary round of selection

The selection procedure allows for the recovery of the desired ancient DNA product and separation from the probe. This procedure was performed upon completion of the 50 hour hybridization incubation period. 100µl (1 mg) of Dynabeads M-280 Streptavidin beads (Dynal, Cat# 112.05) was pipetted into a 1.5 µl eppendorf tube. The tube was placed into a magnetic holder to allow the beads to concentrate at the back of the tube for 1 minute. The buffer was removed from the tube without disturbing the beads. The tube was removed from the holder and the beads were resuspended in 100 µl of TEN buffer (10mM Tris-HCl, pH 7.5, 1mM EDTA, 1M NaCl) with gentle mixing. The tube

was replaced in the holder for 1 minute followed by removal of buffer. This wash step was repeated 2 more times for a total of 3 washes with TEN buffer.

After the third wash the mixture was re-suspended once again in 100 µl of TEN buffer. The hybridization mixture was removed from the water bath and the mixture pipetted carefully into the TEN re-suspended beads without transferring any mineral oil. This mixture was gently vortexed and incubated at room temperature for 20 minutes, and vortexed every 5 minutes to keep the beads suspended. During this incubation 1 Centri-spin column (Princeton Separations, CS-200) per sample was prepared by following the manufacturer's instructions, and the wash buffer II Wash Buffer II (0.1X SSC, 0.1% SDS) solution was prewarmed in the water bath to a temperature of 60˚C.

Following incubation, the tube was placed back into the magnetic holder and the supernatant solution was removed. 1ml of wash buffer I Wash Buffer I (1X SSC, 0.1% SDS) was added to the tube to resuspend the Dynabeads and allowed to incubate for 15 minutes at room temperature. The tube was placed back in the holder for one minute, followed by removal of the buffer. This wash step along with the incubation was repeated 1 more time.

The beads were resuspended in 1ml of warm (60˚C) wash buffer II and incubated at 65˚C for 15 minutes in a water bath. The sample was placed into the magnetic holder and the buffer removed. This incubation and wash step were repeated 2 more times for a total of 3 washes with wash buffer II.

After removal of the final aliquot of wash buffer II, the beads were resuspended in 25µl of freshly prepared 0.1N NaOH and incubated at room temperature for 20 minutes, with gentle vortexing every 5 minutes. This incubation chemically denatures the DNA and separates the captured museum DNA from the bead-bound biotin-labeled probe. The tube was placed back into the magnetic holder and allowed to incubate for 1 minute to allow the beads to concentrate. The supernatant was removed and transferred to another fresh tube containing 25 µl of Tris-HCl, pH 7.5 to neutralize the solution. The neutralized solution was gently vortexed and passed through the prepared Centri-spin 20 column. The samples were labeled "primary (1˚) selected mtDNA".

Primary selection amplification

Four 50µl PCR reactions were set up for each primary selected DNA template, using 10µl of 1˚ selected DNA in each reaction. For a 1X reaction, ORM-28 was used as the primer (0.18 µg), Invitrogen Platinum Taq (0.50 µl), Platinum Taq 10X buffer without $MgCl_2$ (5.00 µl), 50mM $MgCl_2$ (1.50 µl), and 4.00 µl of 10mM dNTP mixture (40.0mM.) The PCR profile used a 1 minute hot start at 95˚C, and 35 cycles of denaturing for 15 seconds at 94˚C, annealing for 20 seconds at 58˚C, extension for 1 minute at 72˚C, followed by a 5 minute final extension at 72˚C. These were labeled "1˚ selected amplified mtDNA." The PCR products were purified with Montage filters (Millipore, Cat# UFC7PCR50) to remove any unincorporated primers and dNTPs. The tubes were labeled "1˚ selected amplified purified mtDNA." The products were resolved on a 1% agarose gel to examine the size-distribution of selected DNA fragments.

Second round of selection and amplification

A second round of DNA selection was carried out, using 1 microgram of the 1˚ selected amplified purified DNA as a template, and 100 ng of the biotin-labeled capture probe, using the same procedure described for the primary selection and amplification. The only difference in the secondary round of selection is the use of 5 µL of template DNA, and 30 rounds of amplification during the PCR profile.

Next generation sequencing, sequence assembly and analysis

The DNA fragments from the secondary selection were resolved on a 1% agarose gel and prepped for DNA sequencing based on the Illumina preparation procedures and based on Illumina manufacturing specifications (Part # 1005361 Rev. B, December 2008). The twelve samples were each given a unique identifier (barcode) by incorporating a unique indexing primer during the final amplification step during library preparation. The sequence libraries were quantified on a Nanodrop device, pooled in equimolar amounts, and sequenced on a single lane of an Illumina flowcell. Initial data filtering and quality checks were done with Illumina software. Sequence data for each individual colugo sample was sorted based on its unique six-base pair index that was incorporated during the library preparation. Sequence data were assembled against the reference mtDNA genome sequence using Maq (http://maq.sourceforge.net/ http://maq.sourceforge.net/). Sequence reads were also evaluated by BLAST (Altschul et al. 1990). Final contigs were assembled in Sequencher to edit ambiguous base calls, evaluate open reading frames, and for exporting alignments in Nexus file format.

Phylogenetic analysis was performed in PAUP (Swofford 2003), and RAXML (Stamatakas 2006) under the maximum likelihood criterion. Confidence levels for nodal support were obtained by bootstrapping (100 heuristic replicates).

CHAPTER III

RESULTS

Probe generation

PCR amplifications of the nineteen primer pairs, shown in Table 2, served as the probe for the hybridization and selection procedures. Homologous segments between museum DNA and probe were able to hybridize with differences in sequence upwards of 10%-15%. This flexibility allows for cross-species hybridization in which the hybridization conditions can be modified based on how divergent the probe is from the sample.

Table 2. Primers used for amplification of probe.

19 Primer Pairs

Start	End	Difference (End - Start of Next Row)	Sequence (start)	Sequence (end)	Product Size
25	1084	237	GCAAGGTACTGAAAATACCAAGATG	TGAAATCTTCCGGGTGTAGG	1060
847	1958	337	CAAAGGAGGATTTAGCAGTAAATTAAG	TGCTAGAGGTGATGTTTTTGG	1112
1621	2681	255	GCCACCAATTAAGATAGCGTTC	CTAACAAGCCCTGCTCTTGG	1061
2426	3809	236	CTCGATGTTGGATCAGGACA	TTCTCAGGAGTGGGTTCGAT	1384
3573	4328	266	CGAGCTTCATACCCACGATT	GGCTAGTTTTTGTCATGTCAGG	756
4062	5340	169	AACCCACGATCAACAGAAGC	AGGGTGAGGTGGCTGAGTAA	1279
5171	6473	154	CTACTTCTCCCGCCTCCAAG	TGTGCTACTACGTAATATGTGTCGTG	1303
6319	7221	246	GCTACACTGCACGGAGGAA	TGGTTTCTACTATTTGGGCATTT	903
6975	8348	571	AAAAACATTACATGACTTCGTCAGA	GGTGTGCCTTGGGGTAGAAG	1374
7777	8951	392	CCACAATGAAATGCCACAAC	TGGAGCTAGGCTTGAGTGGT	1175
8559	9664	800	CACCGTAGCCCTAATCCAAG	ACGTGATGGCCACTAGGAAA	1106
8864	9867	239	ACGATACGGAATAATTCTCTTCA	AATGGGTCGAAACCAGTTGT	1004
9628	10837	287	CCCTTCTCCATAAAATTTTTCC	TTTTGGTAGTCAGAGGTGAAGTC	1210
10550	11697	363	GAAGCAACACTAATCCCAACC	TTGAAAGTAAGAAAGCCATATTTTT	1148
11334	12602	171	CAGCATTCTCCTGATCAAACA	AGTGTGGTGAGGGCACCTA	1269
12431	13869	209	TACACCCGTGACTTCCCTCT	TACTGCCATGGCTATTGAGG	1439
13660	14806	146	GTAGAATCCCCATGAAAATAACC	GGGATTTTGTCTGAGTTTGATG	1147
14660	15919	570	AGACAAAGCCACCCTCACAC	GCATGGCCCTGAAGTAAGAA	1260
15349	16734		CTCCCCAGGACAATCAAGG	GCTTCAGGCCAAAATTCAAA	1386

Shows the coordinates relative to the reference mtDNA genome in GenBank (Acc. No. AJ128849.1), the overlap between fragments, sequence of each forward and reverse primer, and the expected PCR product size.

Amplified adapter ligated museum DNA

Extraction of DNA from ancient museum samples provided variable results of DNA recovery. DNA was recovered from every museum sample, despite the oldest sample being 170 years old. The age of museum samples, however, showed little correlation with quality of DNA recovered (Figure 1). Quality of DNA was measured based on the size and intensity of the DNA smear. The higher the intensity the more product is present, and smaller the fragment size the more degraded the DNA sample (Figure 2). Several samples that were nearly 100 years old showed great quality smears when compared to samples collected more recently. For example, the sample collected in 1903 (index 11) has some of the highest quality DNA even though the sample is now 107 years old; while a sample collected in 1963 (index 7) yielded among the poorest DNA quality. Thus, the quality of DNA recovered from each sample is largely influenced by how the sample was handled and stored during and after collection rather than just its age.

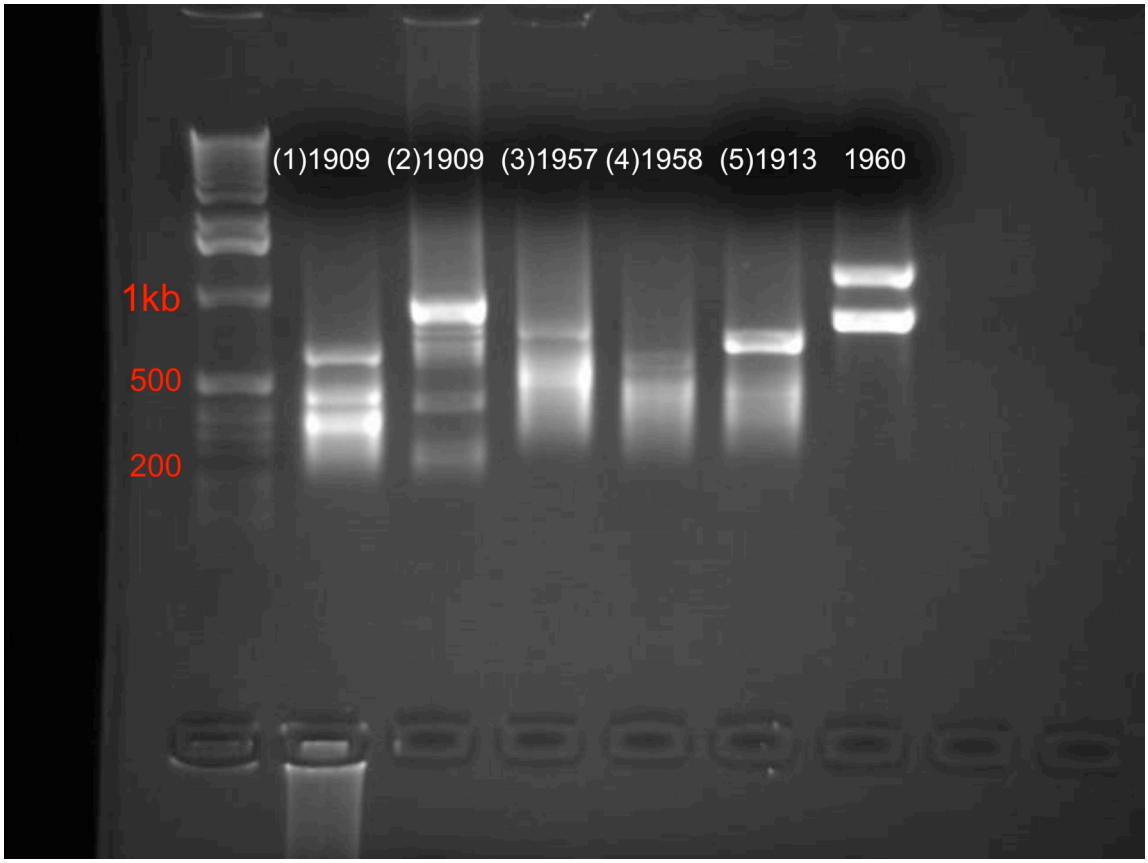

Figure 1. 2˚ Selected amplified mtDNA from museum samples 1-5. These samples are labeled with indexes 1-5 (See Table 1). The index number is listed in parentheses, followed by the sample collection date.

Capture hybridization/mtDNA selection

The distribution and banding patterns of the 2˚ selected amplified museum products are shown in Figures 1 and 3. These have a strong correlation with the percent genome coverage eventually obtained by Illumina DNA sequencing (Table 3). Products in the 200-600 bp range were gel excised, indexed by PCR, pooled and sequenced on an Illumina GAII. Higher intensity bands outside the main fragment smear were excised and sequenced separately with Sanger-based sequencing to reduce bias in sequencing

coverage across fragments in the library. Individuals with uniform smears yielded the highest overall genome coverage percentages because they have a more even distribution and concentration of products. Samples with more banded 2˚ selected amplified products yielded more biased mtDNA genome sequence coverage, with a larger number of gaps. These banded individuals have high concentrations of one sequence rather than having moderate concentrations of many different sequences.

Table 3. Raw sequence data generated from one lane of an Illumina GAII flowcell where 12 individuals were pooled.

	Illumina Results		
Index #	**Total # of Reads**	**Reads Mapped**	**Genome % Coverage**
1	174763	63621	77.38%
2	58444	44019	57.05%
3	123674	89262	95.98%
4	74793	63442	90.21%
5	5178	3146	42.33%
6	1026235	480104	63.98%
7	611655	185891	39.81%
8	391818	284261	42.40%
9	94701	52584	20.64%
10	293723	216895	97.98%
11	133611	110584	89.46%
12	497124	369871	94.71%
Total:	3485719	1963680	
Average:	290476.5833	163640	67.66%

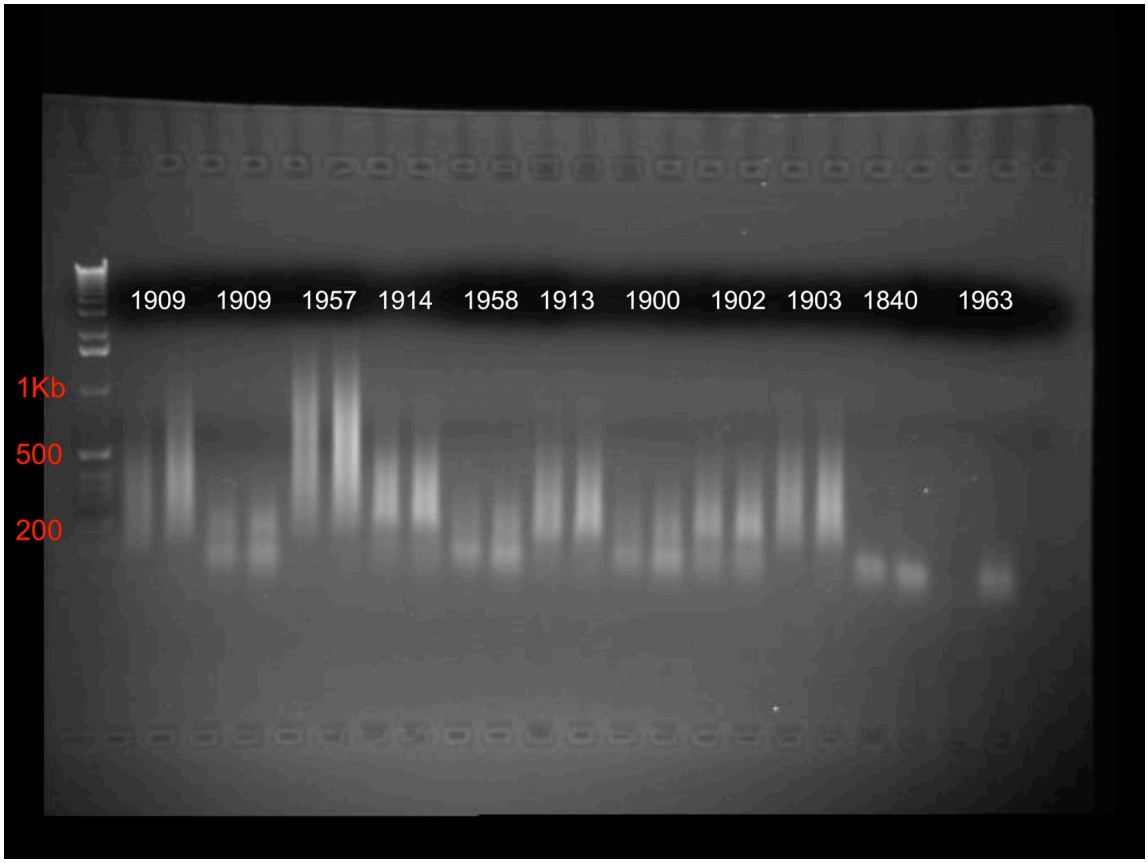

Figure 2. ORM-28 adapter-ligated PCR amplified products of the extracted museum DNA samples. The sample collection dates are shown above each sample, run in duplicate in adjacent lanes of a 1% agarose gel to determine quality of DNA.

Figure 3. 2˚ selected amplified mtDNA from museum samples 7-12. These samples are labeled with indexes 7-12. The index number is in parentheses, followed by the sample collection date.

Preliminary Illumina sequence library analysis using Sanger sequencing

The 2˚ selected amplified products of 143327 and 317119 were produced and evaluated before the remaining samples were processed. Each library was cloned in to the PCR-TOPO Blunt end vector (Invitrogen) and grown on LB+ampicillin plates overnight, following manufacturer's recommendations. 96 colonies from each sequence library were picked into 15 µL of PCR water in a 96-well PCR plate. Two microliters of this template was used in a subsequent PCR reaction using vector-borne universal primers (M13 or T3/T7). PCR products were evaluated on 1% agarose gels and sent to

Agencourt for Sanger DNA sequencing on an ABI-3730 capillary DNA sequencer. The resulting DNA sequences were edited for quality and vector-trimmed in Sequencher (GENECODES, Inc.). The DNA sequences were then assembled relative to the colugo reference mitochondrial DNA genome (AJ428849).

The coverage and distributions of selected products are shown in Figures 4 and 5, relative to a linear depiction of the colugo mitochondrial genome. 92.3% of the high-quality sequences mapped to the reference sequence from sample 143327. Sample 143327 had good genome coverage with a relatively even distribution of DNA fragments (Figure 4). By contrast, 71.9% of the sequences from sample 317119 aligned to the reference genome. Sequence coverage was more biased, with more reads aligning to the highly conserved 12S rRNA region of the genome (the left end of the linear genome) (Figure 5).

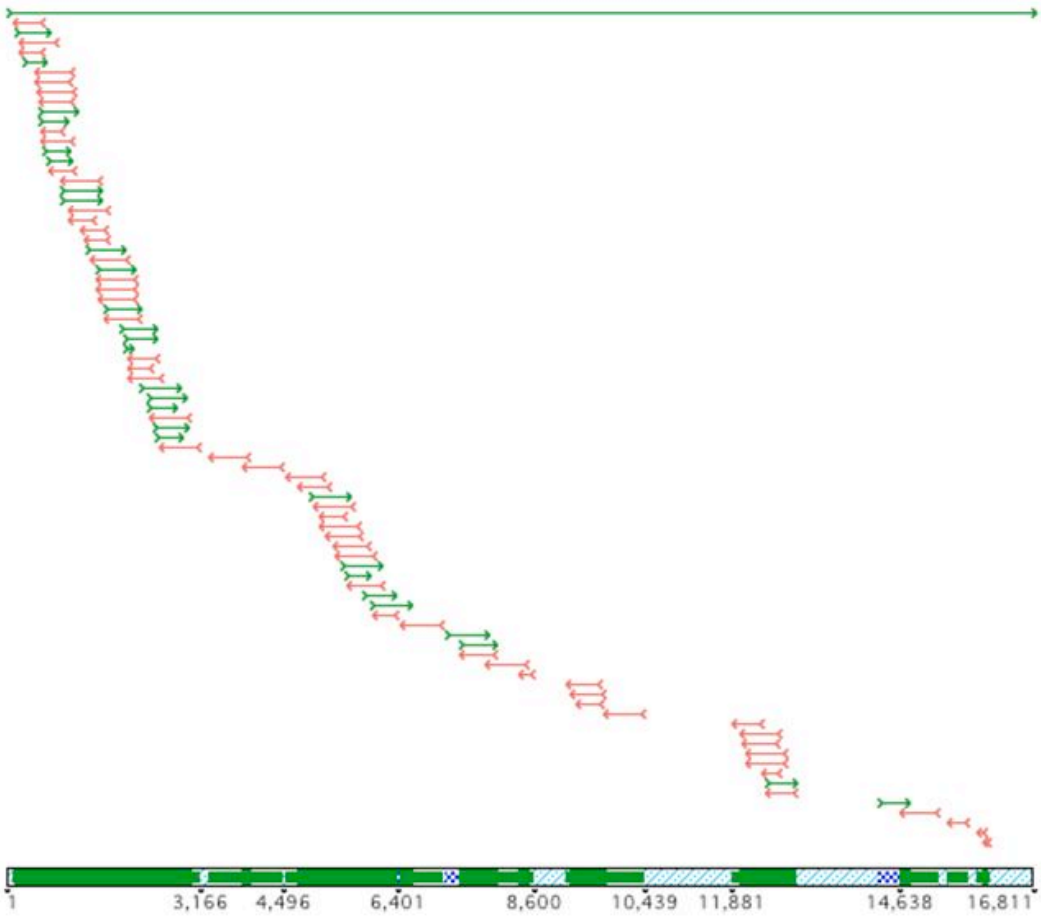

Figure 4. Pilot study and representation of genome coverage for sample 143327. Mitochondrial genome coverage on preliminary data of individual 143327 relative to the reference sequence of a Bornean colugo (GenBank accession number AJ428849) represented by the long green arrow at the top of the figure. The smaller sequences are products after the 2 rounds of selection which were cloned into a plasmid vector. The genome coverage is depicted by the green, blue, and light blue bar at the bottom of the figure. Green regions represent areas of the genome that are covered with more than 1 sequence, the blue checkered areas represent that the reference sequence is covered by 1 sequence, and the light blue means no sequence coverage with respect to the reference.

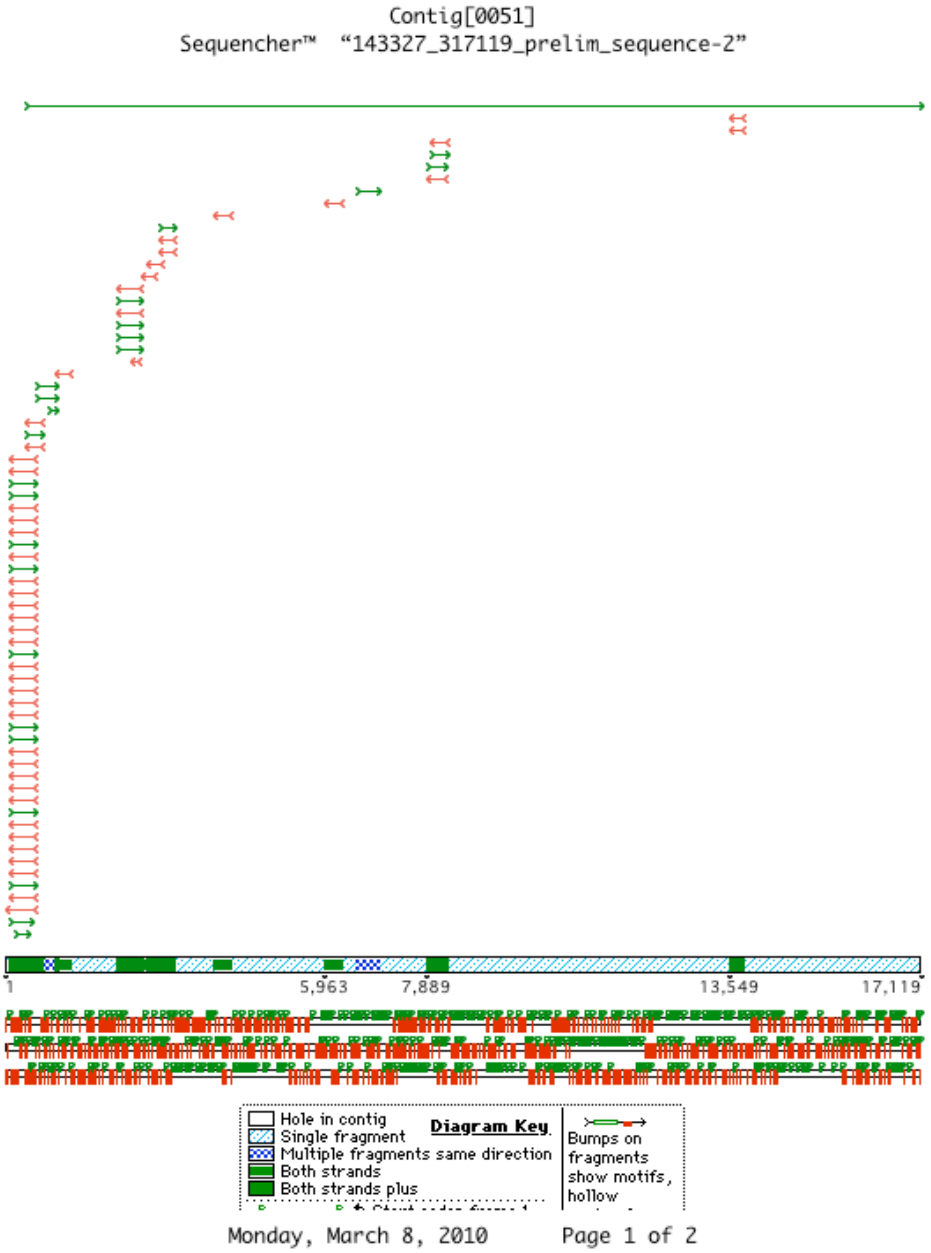

Figure 5. Pilot study and representation of genome coverage for sample 317119. Mitochondrial genome coverage based on preliminary data of individual 317119 relative to the reference sequence of a Bornean colugo (AJ428849), which is represented by the long green arrow at the top of the figure. The smaller sequences are products after the 2 rounds of selection which were cloned into a plasmid vector. The genome coverage is depicted by the green, blue, and light blue bar at the bottom of the figure. Green regions represent areas of the genome that are covered with more than 1 sequence, the blue checkered areas represent that the reference sequence is covered by 1 sequence, and the light blue means no sequence coverage with respect to the reference

Illumina DNA sequencing results

The Illumina Genome Analyzer II is one of several next-generation sequencing platforms that generates DNA sequences with short average reads and high throughput capacity, providing a means to sequence nearly every fragment of DNA in a sample. One lane of a single Illumina GAII flow cell can generate a total of ~220.5 million base pairs from 12 individuals labeled with unique index sequences for identification. The mitochondrial genome of a colugo is 16,748 bps long, so only 200,976 bps of sequence is required to sequence one mitochondrial genome to 1x sequence depth to have complete genome coverage across all 12 samples. This means that if 100% of the sequences were colugo mitochondrial DNA and the concentrations of every base pair in the genome were identical across all twelve individuals, one could achieve ~1097x sequence depth of each base pair. This extreme depth of coverage is often not achieved as every sample is going to have some level of contamination and concentration differences, though it enabled ~99.99% of all base pairs to be accurately called.

Mitochondrial DNA was recovered from every sample submitted for sequencing. Samples that had more bacterial DNA than colugo DNA still provided decent mtDNA genome coverage despite the high level of contamination (Figure 6). Individual 6 was one of the most contaminated samples, where only 37.8% of total sequence reads map to the reference sequence. Despite this high level of contamination, 64% of the mtDNA genome was covered by sequence reads (Table 3).

Figure 6. Schematic depicting the steps undertaken in the preparation of samples submitted for Illumina DNA sequencing. The twelve bars at the bottom of the figure represent the respective genome coverage (in red) for each individual.

Open reading frame analysis

For all colugo samples that generated mtDNA sequence, the thirteen CDS (coding domain sequences) regions of the colugo mitochondrial genome are currently under inspection. Each CDS region should have one continuous open-reading frame in which there is no pre-mature stop codon to arrest translation of the full length protein. Presence of a stop codon that does not map to the correct location is evidence towards selection and incorporation of numts, rather than true colugo mtDNA. Numts, which are former mtDNA that has been translocated into the nuclear genome, have the potential to be captured by the selection procedure and align to the reference sequence. Checking the mitochondrial genome coding regions for numts is essential for authenticating selection results (Triant and DeWoody 2007).

Samples 143327 and 356666, representing indexes 12 and 7 respectively, have been analyzed using only the sequences that align to the reference. In both samples no pre-mature stop codons have been identified. This indicates that the selection procedure has likely been accurate in targeting true mtDNA. In the case of sample 356666 there was only 39.81% genome coverage, which means large areas of some coding regions were absent. The sequence available for analysis was kept in frame and analyzed according to its alignment position on the reference.

Preliminary phylogenetic analysis

The maximum likelihood tree generated by our phylogenetic analysis represents an accurate portrayal of branch length based on genetic divergence, and high bootstrap support values indicate confident branch placement and relationships between individuals from different populations (Figure 7). This preliminary tree does not incorporate all of the sequenced data, which is still being processed to generate the most accurate representation of the mitochondrial genome. The phylogenetic results are concordant with current colugo geographic distributions (Figure 8), as well as the historical biogeography of Southeast Asia. For example, all samples from Peninsular Malaysia are closely related. The Thailand sample, which is also located on the northern Malay Peninsula, shows a close relationship to the southern Malay Peninsula samples. Sumatra, although a separate island, is the next closest landmass to Peninsular Malaysia, and shows little divergence from the Malaysian and Thailand samples. The Natuna Islands sample is the most geographically isolated in the middle of the South China Sea, and very divergent from other samples. The remaining samples from Borneo and Java produce more interesting results. The East Javan sample is very divergent from the West Javan populations, while the reference sequence from Borneo (AJ428849) is more closely related to the Peninsular Malaysian group than the remaining Bornean samples.

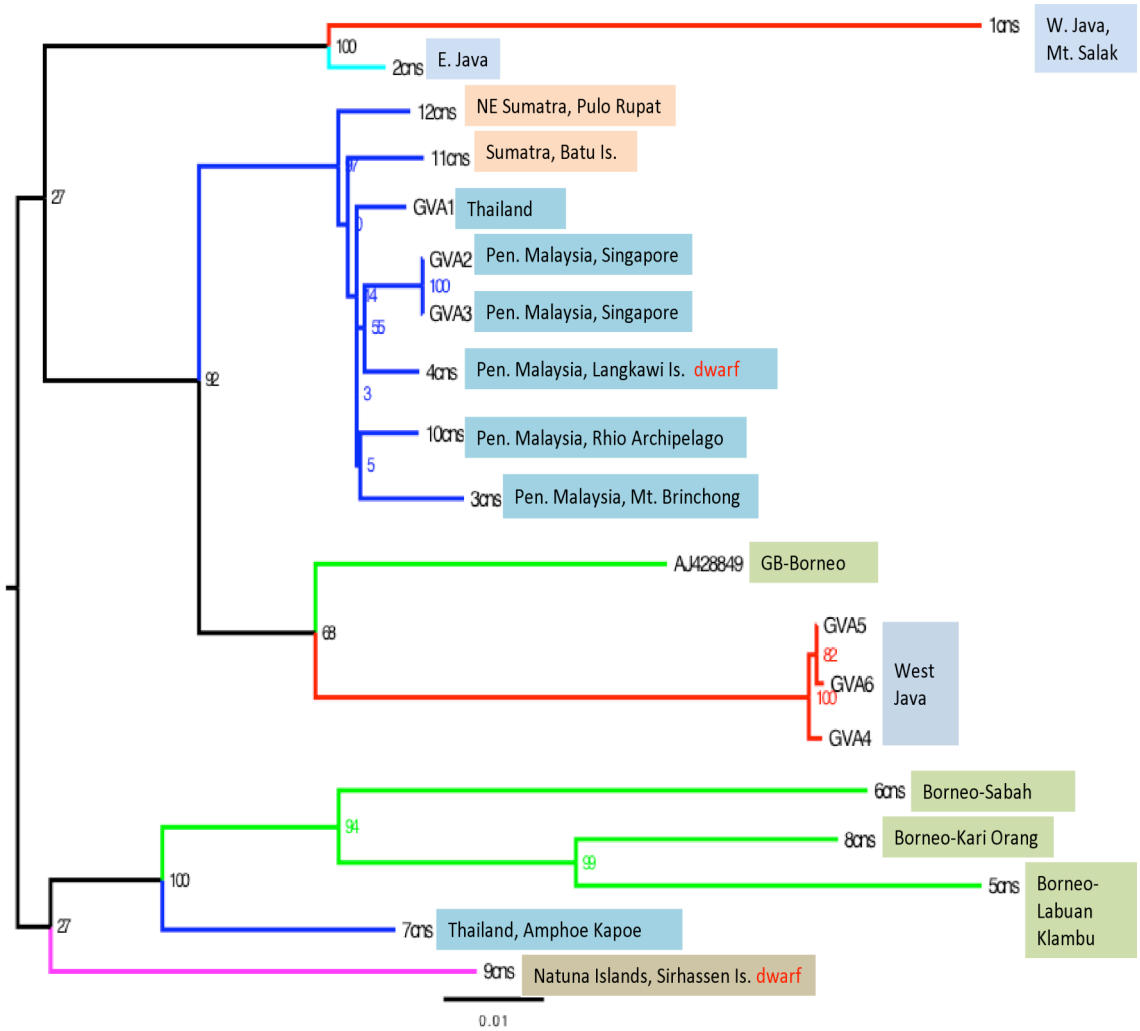

Figure 7. Maximum likelihood phylogenetic tree of colugo mtDNA sequences. Individuals are labeled with their geographic location. Bootstrap values based on 100 heuristic replicates are shown above each branch.

Figure 8. Coordinates for collection locales of each sample, illustrating the biogeographical relationship between individuals.

CHAPTER IV

DISCUSSION AND CONCLUSIONS

Hybridization and genome coverage

Five individuals yielded >90% coverage across the reference mtDNA genome, while others had more restricted coverage (Table 3). Varying amounts of genome coverage may be due to several factors: 1) the initial quality of DNA sample, 2) the level of exogenous DNA contamination in each sample, 3) the divergence between the Javan mtDNA probe and the particular sample, 4) the hybridization temperature conditions 5) multiple rounds of amplification, and 6) the concentration of each segment of probe based on individual average fragment sizes.

The initial quality of a sample, based on the smear of DNA produced following amplification with the adapter ligated primers shown in Figure 1, provides an estimate for the extent of degradation of a given sample, and the potential for efficient hybridization. The initial pool of sequences present in a sample are the only ones available for hybridization and selection, therefore the maximum amount of genome coverage is pre-determined by this initial smear of DNA. A large smear indicates a variety of products available for hybridization, and therefore more potential for genome coverage. Smaller smears tend to give more restricted genome coverage percentages. Long fragments of museum DNA allow for more accurate hybridization to the biotin labeled probe due to more homologous base pairs available for compilation. This yields

more reliable hyrbidization from longer fragments of museum DNA than shorter fragments.

Though the initial DNA distribution can give a good indication as to how much DNA may be available for hybridization, there is no way of knowing whether or not it simply represents bacterial contamination, or DNA from the species of interest. So the DNA that one observes initially could be largely bacterial, or even human, contamination depending on how the sample was handled and stored. The more contamination in a sample gives the hybridization process a larger chance to select for incorrect segments of DNA. Sequences preserved throughout evolutionary time are capable of hybridizing to the probe through conserved regions such as the 12S ribosomal RNA region of the mitochondrial genome. This situation is analagous to designing primers according to one species' DNA sequence and using these primers to amplify DNA from another, related species. This use of cross-species, or heterologous, DNA segments as primers assumes that the primers will hybridize to the other species DNA. The more incorrect DNA hybridized, the more probe is wasted on selecting extraneous DNA, making less probe available for selecting the DNA of interest and more incorrect DNA recovered after the selection procedure. Limiting the amount of probe available for hybridization lowers the amount of DNA able to be recovered and therefore lowers the potential for genome coverage.

One method to mitigate these potential biases is to alter the hybridization conditions according to the relative levels of divergence between the biotin-labeled probe and the target museum DNA of interest. Sequences with small divergence levels readily hybridize together, even under stringent conditions. High incubation temperatures (65˚C) create a stringent environment for the most analogous sequences to hybridize before the more divergent sequences have an opportunity to hybridize to the probe. This has the effect of selecting for sequences that are most likely of colugo mitochondrial DNA origin in each museum sample. Once the most similar sequences are hybridized, slowly lowering the incubation temperatures should take place to capture more divergent sequences. These lower temperatures (~55-60˚C, this varies on how divergent the sequences really are or hypothesized to be), provides flexibility for hybridizing very divergent individuals. The results presented here, where we were able to hybridize sequences upwards of 10-15% divergence from the target probe, provides evidence that cross-species capture hybridization is completely feasible and may be routinely applied for other groups of organisms.

However, lowering the temperature during hybridization does have drawbacks. Unwanted DNA sequences from bacteria, repetitive DNA sequences which are abundant in eukaryotic genomes, or other contaminating DNAs may be recovered. Availability of a reference sequence from your organism of interest, coupled with next generation sequence technologies and BLAST analysis, can largely identify and remove unwanted sequences. Next generation sequencing in particular has such high throughput capacity

that it can feasibly sequence every DNA fragment in a sample library. In this study, a few individuals submitted for next-generation sequencing yielded mostly contaminanting DNA sequences that didn't align to the reference genome. Even in these cases we were able to distinguish true colugo mitochondrial DNA through comparison with the BLAST database and alignment to the reference sequence.

Multiple rounds of amplification preferentially amplifies the more concentrated products in a sample. The more rounds of amplification the more narrow the range of products becomes. Products with very low concentrations simply get out-competed for primer hybridization by products of high concentrations during PCR. Products with low concentrations can be lost during PCR cleanup procedures; although it is more likely that they are lost because relative to the higher concentration products, they become exponentially less abundant following subsequent PCR amplification steps (as the low concentration products are selected against during amplification). Therefore it is advisable to minimize the number of rounds of amplification during this procedure in order to provide the widest distribution of products, and therefore the best genome coverage obtainable.

Genome coverage is also dependent on how the probe is made and how fragments are pooled together. Nineteen primer pairs were constructed for probe preparation, and the PCR products were each around 1 kb in length. If all fragments were exactly 1 kb then one could simply pool the 19 PCR products based on their concentrations. Since these

fragments are of varying length both the concentration and fragment length must be considered when pooling the products for the probe. Longer fragments have higher concentrations with fewer strands of DNA in a sample due to concentration being measured in ng/µL (nanograms per microliter) rather than (number of fragments)/µL. In contrast shorter fragments have lower concentrations with more fragments in the sample. The objective is to get the same number of base pairs contributing to hybridization from each PCR product pooled together. This will provide a more uniform selection of museum DNA during hybridization. If the numbers of base pairs contributing to hybridization vary for different portions of the mitochondrial genome then you will hybridize certain sections of the genome preferentially over others. This subjects the less hybridized portions of the genome to non-preferential amplification procedures and potential for subsequent loss of DNA and coverage for that portion of the genome.

Development of the probe for completely equal numbers of base pairs is not always the most efficient method for equal hybridization. This varies depending on how much DNA of each portion of the genome is present in the museum sample and how divergent the sequences are at different portions of the genome. Areas of the mitochondrial genome are more conserved than others and others more variable between individuals. The 12S ribosomal RNA region of the genome is very conserved among different populations while the hyper-variable region (which consists of various tRNA genes) can vary largely even between individuals in a single population. Therefore the 12S region, which is conserved, will readily hybridize and be much easier to recover in high

concentrations than the hyper-variable region. To combat this, lowering the concentration of base pairs for the 12S ribosomal RNA region and/or raise the concentration of base pairs of the hyper-variable region in the probe, as well as having the hybridization temperature lower for a longer time will allow the more divergent samples to hybridize. Another problem is that the concentration of each base pair in the genome for museum samples is completely unknown prior to beginning the hybridization. This has a major impact on obtaining equal concentrations of base pairs for the genome. Some areas of the genome could be completely absent, or these areas could have widely varying concentrations. With concentrations unknown initially it is hard to obtain equal concentrations of museum DNA after hybridization, selection, and amplification. The objective of the hybridization is to obtain equal concentrations of every base pair throughout the genome after completion of the hybridization, selection, and amplification steps. Many variables and conditions need to be taken into account before attempting the hybridization, as it is what makes up the potential for equal or full genome coverage.

Phylogenetics and biogeography

The drop in sea level during the Pleistocene, along with other dynamic environmental changes created four marked subsections of the Southeast Asian archipelago, which established many isolated regions for speciation events to occur. The sea level drop created many major river systems as well as environmental changes to a dry savanna like state (Harrison et al. 2006). Being arboreal mammals, colugos cannot easily survive in a

savannah like ecosystem, nor can they traverse large rivers safely (Lim 2007). For colugos, this means that the drop in sea level did not simply create open access to all different landmasses throughout the archipelago. Thus, the four different subsections that sustained refugia for arboreal mammals were (1) central and northern Borneo; (2) Malay Peninsula including Sumatra; (3) Mentawai Islands; and (4) Western Java (Harrison et al. 2006). A belt of dry woodland and savanna extended from southern and eastern Borneo all the way down to Eastern Java (Harrison et al. 2006). The isolation of Western Java provides an explanation for the observed genetic divergence from the adjacent East Javan samples in our phylogenetic tree in Figure 7. Borneo itself was also subdivided into two distinct ecological regions: the tropical refugia of the north and west, and the contrasting dry savannah of the southeast. Borneo is a very mountainous landmass with mountain ranges running from the center to the northeastern tip of the island. Borneo also has many major river systems running throughout the southern portion of the island. This plethora of geographical isolating mechanisms would have provided many opportunities for unique isolating events, and supports the unexpected phylogenetic distribution and deep divergence of the samples studied here within Borneo.

Under the current taxonomy, individuals in this study are classified as one species, *Galeopterus variegatus*, the Sunda Colugo. An analysis of genetic divergence between new and published mitochondrial sequences was performed in a maximum likelihood phylogenetic framework (Figure 7). Figure 7 indicates very large divergences between

individuals, and potentially several distinct species. Results of this study show different populations which are currently classified as one species, distinguished by large amounts of previously unknown genetic divergence. Though molecular divergence dates were not performed on this dataset, a previous study by Janecka et al. (2008) revealed potential species level distinctions between GVA 4, 5, and 6 from West Java and GVA 1, 2, and 3 from the Malay Peninsula which have ~5.4 million years of divergence between the geographical locations. Referencing Figure 7, one can see that there is upwards of 3 times the genetic divergence between those populations from Janecka et al. (2008) and the new locales sampled in this study. Alone, the five million years of divergence between Javan and mainland populations were considered equivalent to species level distinctions. Thus the threefold observed divergence illustrated by figure 7, leaves little doubt that there likely are many potential species level distinctions within the Sunda colugo populations. What other locals throughout Sundaland could harbor similar, or even greater, levels of this cryptic genetic divergence? What of the unsampled Southeast Asian mainland, such as northern Thailand, Laos or Vietnam? The Philippine colugo, *Cynocephalus volans*, has also has many isolated populations interspersed throughout multiple islands that may have been ideal for founder events as well. Though this study provides reasonable coverage of different geographical locations throughout the Southeast Asian archipelago, it is by no means a finished study. Many other studies can efficiently stem out of this analysis with application of hybridization and selection procedures, taking advantage of simple sampling methods by

utilization of extensive and untapped museum samples, and the promise of hidden genetic divergence throughout the Southeast Asian archipelago.

REFERENCES

Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ (1990) Basic local alignment tool. *Journal of Molecular Biology* **215**: 403–410.

Harrison T, Krigbaum J, Manser J. 2006. Primate biogeography and ecology on the Sunda Shelf Islands: A paleontological and zooarchaeological perspective. In *Primate Biogeography* (eds. Lehman SM, Fleagle JG), pp. 331–372. Springer Science & Business Media, New York.

Janečka, JE, Miller W, Pringle TH, Wiens F, Zitzmann A, Helgen KM, Springer MS, Murphy WJ. 2007. Molecular and genomic data identify the closest living relative of primates. *Science* **318:** 792–794.

Janečka JE, Helgen KM, Lim NT-L, Baba M, Izawa M, Boeadi, Murphy WJ. 2008. Evidence for multiple species of Sunda colugos. *Current Biology* **18**: R1001–R1002.

Lim, NT-L. 2007. *Colugo: The Flying Lemur of South-east Asia*. Draco Publishing and Distribution, Singapore.

Lomolino MV, Riddle BR, Brown JH. 2006. *Biogeography*, 3rd edition. Sinauer Associates, Sunderland, Massachusetts.

Millar CD, Huynen L, Subramanian S, Mohandesan E, Lamber DM. 2008. New developments in ancient genomics. *Trends Ecol. Evol.* **23**: 386–393.

Peterson AS. 1998. Direct cDNA selection. In *Genome Analysis: A Laboratory Manual* (eds. Birren B, Green ED, Klapholz S, Myers RM, Roskams J.), pp. 159–171. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Ruggeri N, Etterson M. 1998. The first records of colugo (*Cynocephalus variegatus*) from the Lao PDR. *Mammalia* **62**: 450–451.

Schipper J, Chanson JS, Chiozza F, Cox NA, Hoffmann M, Katariya V, Lamoreux J, Rodrigues AS, Stuart SN, Temple HJ, et al. 2008. The status of the World's land and marine mammals: Diversity, threat, and knowledge. *Science* **322**: 255–230.

Stafford BJ, Szalay FS. 2006. Craniodental functional morphology and taxonomy of dermopterans. *Journal of Mammology* **81:** 360–385

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analysis with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

Swofford DL. 2003. *PAUP*: Phylogenetic Analysis Using Parsimony* (*and other methods). Sinauer Associates, Sunderland, Massachusetts.

Templeton AR. 2006. *Population Genetics and Microevolutionary Theory.* John Wiley & Sons, Hoboken, New Jersey.

Triant DA, DeWoody JA. 2007. The occurrence, detection, and aviodance of mitochondrial DNA translocations in mammalian systematics and phylogeography. *Journal of Mammology* **88**: 908–920.

Woodruff DS. 2003. Neogene marine transgressions, paleogeography and biogeographic transitions onthe Thai–Malay Peninsula. *Journal of Biogeography* **30**: 551–567.

CONTACT INFORMATION

Name:	Victor C. Mason
Professional Address:	Department of Veterinary Integrative Biosciences TAMU 4458, Texas A&M University College Station, TX 77843-4458
Email Address:	victorcmason@yahoo.com
Education:	B.S., Biomedical Science, Texas A&M University, May 2010 Undergraduate Research Scholar