REASSESSING COLUGO PHYLOGENY, TAXONOMY, AND BIOGEOGRAPHY BY GENOME WIDE COMPARISONS AND DNA CAPTURE HYBRIDIZATION FROM MUSEUM SPECIMENS

A Dissertation

by

VICTOR CHRISTIAN MASON

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	William J Murphy
Committee Members,	Kristofer M Helgen
	Paul Samollow
	Tiffani Williams
	James Cai
Head of Department,	Dorothy Shippen

May 2016

Major Subject: Genetics

Copyright 2016 Victor Christian Mason

ABSTRACT

The ability to uncover the phylogenetic history of archived museum material with molecular techniques has rapidly improved due to the reduced cost and increased sequence capacity of next-generation sequencing technologies. However it remains difficult to isolate large, orthologous DNA regions across multiple divergent species. Here we describe the use of cross-species DNA capture hybridization techniques and next-generation sequencing to selectively isolate and sequence mitochondrial DNA genomes and nuclear DNA from the degraded DNA of museum specimens, using probes generated from the DNA of an extant species.

Colugos are among the most poorly understood of all living mammals despite their central role in our understanding of higher-level primate relationships. Two described species of these extreme gliders are the sole living members of a unique mammalian order, Dermoptera, distributed throughout Southeast Asia. We generated a draft genome sequence for a Sunda colugo and a reference alignment for the Philippine colugo, and used these to identify colugo-specific enrichment in sensory and musculo-skeletal related genes that likely underlie their nocturnal and gliding adaptations. Phylogenomic analysis and catalogs of rare genomic changes overwhelmingly support the hypothesis that colugos are the sister group to primates (Primatomorpha), to the exclusion of treeshrews. We also captured ~140-kb of orthologous sequence data from colugo museum specimens sampled across their range, and identified deep genetic structure between many geographically isolated populations of the two named species,

consistent with a remarkable increase in diversity. Our results identify conservation units to mitigate future losses of this enigmatic mammalian order.

Examining multiple distantly related mammals we identified a consistent pattern of early diversification between east and west Borneo including colugos, the lesser mouse deer, and pangolins. This strongly parallel biogeographic pattern is not common in mammals and we see no evidence for this pattern in the greater mouse deer. Colugos on West Borneo diverged from those in Indochina in the late Pliocene, however most other mammals across this same geographic region diverged from their common ancestor much more recently in the Pleistocene. Low genetic divergence between colugos on large landmasses and colugos on neighboring islands indicate that past forest distributions in the recent past were recently much larger than present refugial distributions.

ACKNOWLEDGEMENTS

I want to start off by thanking my PI, committee chair, and friend Dr. William Murphy for not only giving me an opportunity to pursue my scientific interests, but also for being the best role-model a student could ask for. He has pushed me to strive for goals I assumed to be out of reach, and has always been available and happy to discuss my sometimes rather ridiculous ideas. Dr. Murphy has always put the well-being of his students first and given myself and others substantial scientific and personal opportunities to develop and grow to be better people. I attribute a large part of what I have achieved and what I want to achieve to him alone.

I would like to thank Dr. Kristofer Helgen for being exceptionally positive, and an enormous facilitator of this dissertation. He traveled around the world with me to sample museum specimens and measure hundreds of colugo skulls used for this dissertation. In addition, his truly encyclopedic knowledge of mammals has generated integral intellectual contributions to our hypotheses and conclusions.

I would like to thank my committee member Dr. Paul Samollow for taking the time to critically think and comment on my dissertation. Dr. Samollow has always understood the true meaning of things. He always has something insightful and uplifting to say and I value all of our conversations. My interactions with both Dr. Murphy and Dr Samollow as an undergraduate student convinced me to pursue a doctoral degree.

I also want to thank my committee member Dr. Tiffani Williams for introducing me to Python which allowed me to apply my own ideas to big data, and for reminding

me that reproducibile results are the only real scientific results. Thank you to my committee member Dr. James Cai for broadening my bioinformatic knowledge and being a part of this entire journey.

Thank you to everyone that has been a part of the Murphy lab over the years, Particularly Dr. Gang Li, Dr. Brian Davis, and M.S. Wesley Brashear. Dr. Li has answered so many of my questions over the years, and been a great friend. Thanks Wesley for being a backboard that I can bounce ideas off of at any moment and being an exceptional friend.

Graduate school has been difficult at times, and my family has stood beside me through everything, my parents Jacque and Richard Mason and, my brothers Austin and Owen Mason. Of course all of my friends and people close to me throughout graduate school have encouraged me, humbled me, and heavily influenced how I see the world. You know who you are and thank you.

Also thank you to the Texas A&M University (TAMU) College of Veterinary Medicine, TAMU Institute for Genome Sciences and Society, the TAMU Genetics Graduate Student Association, the National Human Genome Research Institute (NHGRI), and the National Science Foundation (NSF) for funding various aspects of this project.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

1.1 Colugos

Colugos are arboreal nocturnal mammals that remain poorly understood despite being known to science for centuries (Linnaeus 1758). Colugos have undergone extreme adaptive changes from their scansorial and insectivorous ancestor (Bloch 2007) to glide through high-density forested canopies and survive on a largely folivorous diet. One such adaptation is their patagium, a dermal gliding membrane that interconnects all appendages and digits. The patagium is certainly the colugo's most striking feature and explains why researchers named their order Dermoptera which translates to skin-wing (Beard 1993, Lim 2007). This patagium allows them to glide for long distances however colugos are incapable of true powered flight. Their specialized musculoskeletal morphology has also includes elongated arms, legs, and phalanges which increase surface area of their patagium and reduce weight to aid in gliding (Stafford 1999, Beard 1993). These adaptations enable colugos to glide for greater than 100 meters (m) with a 10:1 horizontal to vertical distance ratio (Lim 2007). By analyzing takeoff and landing kinetics, Byrnes (2008) identified a negative correlation between glide length and landing force. Colugos reorient their bodies vertically immediately before landing to increase lift and drag and reduce velocity prior to impact. Long glides allow colugos to correctly position themselves for landing reducing landing impact and lowering risk of

injury (Byrnes 2008). Gliding for large distances would be advantageous for increased resource acquisition and for escaping predators (Byrnes 2008).

While hanging from trunks or branches of trees colugos can be remarkably well camouflaged (Chasen & Kloss 1929, Lim 2007). Both sexes have highly cryptic dappled spotting patterns for camouflage against flecked tree bark. Coat coloration is sexually dimorphic. Females are usually a highly cryptic grey while males can be a very vibrant rufous (Chasen & Kloss 1929, Lim 2007). These colorations and coat patterns are only visible on the dorsum suggesting that they are adaptations for camouflaging colugos in forested canopies. However, it is also possible that some male coloration has evolved not only for camouflage, but also for an alternative function such as sexual attraction.

It is likely that remaining in the forest canopy as much as possible is beneficial to colugos, as this provides an effective means of movement, continuous access to food, and effective camouflage from predators such as raptorial birds (*Pithecophagga jefferyi* and *Spizaetus cirrhatus*), Sumatran tigers (*Panthera tigris*), and long-tailed macaques (*Macaca fascicularis*) (Lim 2007). Colugos have occasionally been observed on the ground where they move in a series of short inefficient hops, during which they attempt to throw their patagium over obtrusive logs and twigs (Lim 2007). A similar hopping motion is observed when ascending trees (Lim 2007). This type of movement appears very inefficient, likely expending large amounts of energy compared to their generally conservative behaviors of hanging and gliding. This highlights the importance of efficient gliding.

Despite their gliding prowess colugos spend relatively little time moving from tree to tree. Rather colugos spend most of their time hanging from branches and feeding on young leaves (Lim 2007). To retrieve the most nutrients from these leaves colugos have evolved an enlarged cecum (48cm) and a four meter long intestine (Lim 2007). Even still, leaves provide little nutritional value making energy conservation important for colugos. To reduce energy expenditure while hanging colugos have adapted non-retractable claws allowing them to hang with little to no effort (Beard 1993, Bloch 2007, Lim 2007).

Colugos are arboreal specialists, but are also dietary generalists. Wischusen and Richmond (1998) recorded Phillipine colugos foraging from 35 of a total of 76 total species of trees in the study area. Colugos were noted to completely exclude a family of trees (Fagaceae) that is known to have high tannin levels which can inhibit digestion, and favor trees of the Myrtaceae family for food and shelter (Lim 2007). Being able to forage from many different food sources lowers the energy cost for traveling to specialized food sources. For long-term survival colugos require high density evergreen forest canopies (>95% cover) (Lim 2013). Outside of dense forests, colugos are likely to face increased predation, would have no source of food, and no trees would be available for efficient movement, making it highly unlikely that colugos could survive or disperse across such environments. Therefore, colugos are restricted to forested environments and require forested corridors for dispersal, and are incapable of moving through savanna or alpine environments (Chasen & Kloss 1929). Despite these limitations, colugos are one of the most widely distributed mammals across the Southeast Asian mainland and exist

on more than fifty islands currently isolated by large expanses of shallow seas across the Southeast Asian Archipelago and Southern Philippines (Lim 2007). Colugos' inability to fly makes them unable to disperse overwater which is consistent with their absence on deep water islands outside of the Sundaic or Philippine island systems (Chasen & Kloss 1929). This inability to disperse across water suggests that past forested distributions were much more extensive than at present, and that currently isolated islands were once interconnected by forested habitat during periods of low sea levels (Cannon et al. 2009, Woodruff et al. 2011, Raes et al. 2014, de Bruyn et al. 2014).

1.2 Current and Historical Taxonomy of Dermoptera

Current taxonomic classification place colugos within order Dermoptera (Illiger 1811), family Cynocephalidae (Simpson 1945), with two monotypic genera *G. variegatus* (Audebert 1799) and *C. volans* (Linnaeus 1758) (Thomas 1908, Wilson & Reeder 2005, Jackson & Thorington 2012). Four subspecies of the Sunda colugo are supported by craniodental morphometric variation, *G. variegatus variegatus* representing colugos from Java, *G. variegatus borneanus* representing colugos from Borneo and surrounding islands, *G. variegatus peninsulae* representing colugos from Peninsular Malaysia, the surrounding islands, and the Southeast Asian mainland, and *G. variegatus temminckii* representing Sumatra and surrounding islands (Stafford & Szalay 2000) (Table 1.1). Although twenty distinct subspecies classifications were proposed for the Sunda colugo in the early twentieth century, these are currently synonymized within the four accepted Sunda colugo subspecies (Table 1.1). Additional potentially distinct

subspecies classifications has been noted for dwarf populations of colugos within Sundaland (Stafford & Szalay 2000, Jackson & Thorington 2012).

Table 1.1. List of currently accepted and historically proposed Sunda colugo subspecies. Nomenclature: A = currently accepted, P = previously proposed. Subspecies names and locations derived from Chasen & Kloss (1929), and citations for original descriptions derived from Jackson & Thorington (2012). Many subspecies were proposed under the genus name of *Galeopithecus*. In these cases we have updated the genus to its current name, *Galeopterus*.

#	A/p	Subspecies	Synonymized with:	Geographic Location	Original Description
1	A	*G. v. variegatus*	n/a	Java	Geoff 1829
2	A	*G. v. peninsulae*	n/a	Malay States	Thomas 1908
3	A	*G. v. temmincki*	n/a	Sumatra	Waterhouse 1839
4	A	*G. v. borneanus*	n/a	SE. Borneo	Lyon 1911
5	P	*G. v. abotti*	*borneanus*	Penebangan Indonesia, W. Borneo	Lyon 1911
6	P	*G. v. gracilis*	*borneanus*	Pulau Serasan (or Sirhassen)	Miller 1903
7	P	*G. v. lechei*	*borneanus*	Central E. Borneo	Gyldenstolpe 1920
8	P	*G. v. hantu*	*borneanus*	North Sarawak, Borneo	Cabrera 1924
9	P	*G. v. lautensis*	*borneanus*	Pulo Laut Indonesia, SE. Borneo	Lyon 1911
10	P	*G. v. natunae*	*borneanus*	Pulau Bunguran	Miller 1903
11	P	*G. v. perhentianus*	*peninsulae*	East Perhentian Island	Chasen & Kloss 1929
12	P	*G. v. chombolis*	*peninsulae*	Pulau Chombol, Rhio Archipelago	Lyon 1909
13	P	*G. v. taylori*	*peninsulae*	Pulau Tiomon	Thomas 1908
14	P	*G. v. aoris*	*peninsulae*	Pulau Aur (or Aor)	Miller 1903
15	P	*G. v. terutaus*	*peninsulae*	Pulau Terutau	Chasen & Kloss 1929
16	P	*G. v. pumilus*	*peninsulae*	Pulau Adang	Miller 1903
17	P	*G. v. undatus*	*variegatus*	Sumatra? / Java?	Wagner 1839

Table 1.1 Continued.

#	A/p	Subspecies	Synonymized with:	Geographic Location	Original Description
18	**P**	***G. v. saturatus***	***temmincki***	**Pulau Tanah bala, Batu Islands**	**Miller 1903**
19	P	*G. v. tellonis*	*temmincki*	Pulau Tello, Batu Islands	Lyone 1908
20	P	*G. v. tuancus*	*temmincki*	Pulau Tuangku, Banjak Islands	Miller 1903

Species level taxonomic relationships between different colugos has also been inconsistent over time. Various numbers of species have been classified using different species concepts. Here we define four species concepts as follows: 1) Genetic Species Concept **–** GSC – "a group of genetically compatible interbreeding natural populations that [are] genetically isolated from other such groups" (Baker & Bradley 2006); 2) Biological Species Concept **–** BSC – "a group of interbreeding natural populations that is reproductively isolated from other such groups" (Mayr 1942); 3) Morphological Species Concept – MSC – a group of organisms that are of the same species by physical or morphological similarity; 4) Phylogenetic Species Concept **–** PSC – "the smallest population or group of populations within which there is a parental pattern of ancestry and descent and which is diagnosable by unique combinations of character states" (Cracraft 1997) (Other definitions presented in Coyne & Orr 2004). Many species of colugos were defined in the early twentieth century based upon the MSC (Lyon 1908, 1911, Miller 1900, 1903, 1906, Thomas 1908, Stafford 1999). The number of species increased rapidly as many species were defined largely on the basis of body size variation (i.e. dwarfism) or fixation of coat coloration on specific islands. This continued until a maximum of 25 species was defined by Cabrera in (1925). However, twenty of

these species were reduced to subspecific rank by Chasen & Kloss (1929) (Table 1.1). Thomas (1908) was the first to describe *Galeopterus* and *Cynocephalus* as distinct genera, however for much of the twentieth century colugos were still considered two species of the same genus (*Cynocephalus*) because Simpson (1945) referred to them as such. It was not until Stafford & Szalay (2000) that the two species of colugos were accurately described as two distinct monotypic genera, and the generic names of *Galeopterus* and *Cynocephalus* were reintroduced to the literature. More recently, preliminary genetic data (Janecka et al. 2008, Mason et al. 2011) suggests that a larger number of colugo species should be recognized to accurately represent levels of genetic divergence observed across and within insular Sundaic populations, with potential species level differences identified between Peninsular Malaysia, Borneo, and Java based upon the PSC / GSC. This could mean that several of the currently recognized subspecies that were previously classified as species based on the MSC, should be reestablished as full species based on the GSC.

1.3 Historical Interordinal Relationships of Dermoptera

Linnaeus (1758) first described the Philippine colugo as *Lemur volans*, while the Sunda colugo was not named until 1799 (Audebert 1799). The phylogenetic placement of colugos has been controversial since their initial description. Over the years, many relationships and nomenclatures have been proposed (Boddaert 1768, Rafinesque 1815, Thomas 1908a, Jackson 2012). One proposal suggested that bats, primates, tree-shrews, elephant-shrews, and colugos were monophyletic within the taxonomic order Archonta (Gregory 1910). Another termed "the flying primate hypothesis" started with Linnaeus

in 1758 and was expanded upon by Smith in 1980. The flying primate hypothesis states that colugos are the basal lineage to primates and megabats where colugos represent a transitional gliding form prior to the formation of flying megabats. Similarly the supraorder Volitantia (Leche 1886) was revived by Novacek & Wyss (1986) to unite dermopterans and bats based upon their similar interdigital patagia. Colugos were also thought to represent one of several different mammalian orders that separate microbats from megabats based on neurological and various morphological characters (Pettigrew et al. 1989, 1995, 2008, Maseko 2007). These hypotheses have been challenged by conclusions based upon molecular sequence data (Ammerman & Hillis 1992, Murphy et al. 2001, Schmitz 2002, Gunnel & Simmons 2005, Teeling et al. 2005, Meredith et al. 2011, O'Leary 2013). Arnason et al. (2002) found colugos to group within primates as the sister lineage to catarrhine primates (Old and New world monkeys and apes) based on mtDNA evidence. Primates paraphyly was also recovered from a reanalysis of mtDNA evidence by Schmitz (2002); however the similar amino acid and nucleotide compositions of simian and colugo mtDNA genomes was determined to be misleading inasmuch as nuclear DNA results based on the insertion patterns of retroposon elements supported colugos to be sister to Primates (Primatomorpha) (Schmitz 2002).

Primatomorpha was first described by Beard (1991, 1993), who suggested that extinct paromomyids and micromomyids (primitive primate-like euarchontans) were closely related to Dermoptera based on morphological characters indicative of gliding in colugos, such as elongated intermediate phalanges. More recent molecular studies have supported Primatomorpha with combined mitochondrial and nuclear DNA phylogenies,

as well as and protein coding gene insertions and deletions (INDELs) (Janecka et al. 2007, Meredith et al. 2011). However, Bloch et al. (2007) reported strong cladistic support for the monophyly of colugos and treeshrews (Sundatheria) as the sister lineage to primates based on morphological characters when bats were excluded from morphological comparisons. In addition, the morphological link of mitten-gliding between Dermoptera and extinct lineages of Paromomyidae and Micromomyidae proposed by Beard (1991,1993) was refuted by Bloch et al. (2007) based on new specimens of Paromomyidae and Micromomyidae that completely lacked characters involved in mitten-gliding or quadrupedal suspensory adaptations. More recently O'Leary et al. (2013) also placed Sundatheria as sister to primates based on 69 phenomic characters. In summary, the placement of order Dermoptera within the mammalian tree has remained contentious and requires further data to resolve the current phylogentic position of Dermoptera within Mammalia.

1.4 Biogeography

Colugos have one of the most expansive distributions of any mammal throughout the geologically dynamic Philippine and Sunda shelves. This widespread distribution and limitation to forested habitats make colugos a valuable biomarker to study past forest distributions and correlations between biology and geography. Both Sundaland and the Southern Philippines are well known for harboring extensive biological diversity and have been designated as evolutionary hotspots (Myers 2000, de Bruyn et al. 2014). Sundaland is a large peninsula protruding south of the Himalayas and extending from Laos south to Java. Much of Sundaland is currently flooded by shallow seas that separate

the many islands in the Southeast Asian archipelago. However, throughout the Miocene (<23-5.3 Mya) dramatic geological and environmental changes shaped most of the modern structure of Sundaland through highland formation, growth of volcanic island chains such as the Sulu archipelago (~10Mya), and flooding of the Sunda shelf ~5 Mya, which isolated mainland Indochina from the rest of the sundaic islands.

Changes in global climate have resulted in fluctuating sea-levels throughout time; however since the late Pliocene (~3.2Mya) global climate fluctuated with increased frequency and intensity by cycling between long glacial periods and intermittent, and comparably short, interglacial periods (Uba 2007, Cannon et al. 2009, Woodruff 2010, Lohmann et al. 2011, de Bruyn et al. 2014). Pleistocene glacial periods were characterized by low global sea levels due to lowered global temperatures and extensive glacial formation. This generally resulted in a drier climate because of the reduced overland surface-area of water for evaporation (Hope 2007, Lohmann et al. 2011). Interglacial periods however were hot and wet, hence glaciers melted as a result of rising temperatures, and the increase in the surface area of water allows increased evaporation and precipitation (Morley 2000). Global sea levels rose and fell over 120m between interglacial and glacial maxima (Voris 2000, Sathiamurthy & Voris 2006). As the climate cooled glaciers formed, sea levels fell, and the shallow seas of the Sunda shelf disappeared. These lowered sea levels resulted in the merger of many islands by subaerial corridors that could facilitate dispersal of taxa between islands (Voris 2006). Even minor drops in sea level of -35m below present levels connected major islands of Sumatra and Borneo (Sathiamurthy & Voris 2006). By contrast, during high sea levels,

island populations became isolated from the mainland or fragmented into additional, smaller islands. The intensity of glacial and interglacial periods varied, and some glacial periods lowered sea levels in Sundaland by only -40m, while others resulted in a glacial maxima scenario (-120m). It has been proposed that during glacial maxima a north-south savannah corridor and/or deciduous (seasonal) forest extended from eastern Peninsular Malaysia through the exposed South China Sea between Sumatra and Borneo, and that this curved east through the Java Sea and extended south through eastern Java and then back northward through southern and eastern Borneo (Heaney 1991, Meijaard 2003, Bird 2005, Harrison 2006). This scenario is in agreement with the likely dry environmental conditions and coarse sandy soils of the Sunda shelf, which are thought to hinder propagation of forests (Slik 2011). A savanna would be impenetrable for colugos which require evergreen forests for long-term survival. However, many forest dependent mammalian and avian taxa have exchanged maternally inherited mtDNA within the Pleistocene between Sumatra / Peninsular Malaysia and Borneo suggesting that this savanna corridor was not continuous during at least some glacial periods (Leonard 2015), and some reconstructions argue for the presence of a dipterocarp forested connection between Sumatra/Peninsular Malaysia and Borneo as early as the last glacial maximum (~0.015Mya) (Raes 2014). It should be noted that ‘forest dependent’ includes terrestrial taxa like rodents, or taxa like birds which are capable of flight, and such taxa might be more capable of dispersal across alternative environments when compared to the highly limited dispersal of colugos.

1.5 Application of Capture Hybridization Techniques to Museum Specimens

To begin to understand the genetic variation within and between colugo populations a larger sampling of tissues from colugos across the mainland and many islands of the Sundaic and Philippine archipelagos is needed as previous studies have relied on only a few individuals (Janecka et al. 2008) to represent major geographical regions. However, acquiring fresh tissues from colugos is extrememly problematic because they are nocturnal, arboreal, spread across large geographic distances, and do not survive in captivity (Lim 2007). The logistic difficulties of field sampling and lack of captive populations prompted us to utilize museum specimen collections for obtaining specimens for our genetic studies.

Historical tissues acquired from museums are a potentially rich resource for molecular genetic studies. Samples can be collected quickly and cheaply from a variety of locations, and phenotypes can be characterized and compared between specimens. Unfortunately DNA quality varies greatly among museum specimens, and even different tissues from the same specimen (Binladen et al. 2006, Mason et al. 2011). Next generation sequencing technologies have enabled economical sequencing of large amounts of these degraded DNA fragments, which was previously not possible. Preliminary evidence shows that brain and nasal crusties (dried adherent tissue inside cranial and nasal cavities) provide the most DNA/mg of tissue. Being protected inside the skull these tissues are less prone to cross contamination than skin or hair, and allow for minimally invasive sampling.

Nevertheless, museum DNA quality is unpredictable and DNA extracts are contaminated with exogenous DNA. Moreover, classic orthologous DNA enrichment techniques such as PCR are only capable of recovering ~100bp DNA fragments with possible amplification of contaminating sequences due to the degraded and contaminated DNA pools, making this process laborious, expensive, and inefficient. Capture hybridization procedures offer a more efficient and broadly applicable approach for the analysis of degraded and contaminated DNA pools, and have been successfully utilized to enrich for orthologous DNA sequences from historical museum DNA (Mason et al. 2011, Bi et al. 2013, Hawkins et al. 2015) as well as from ancient DNA >300,000 years old from extinct cave bears (*Ursus deningeri*) where DNA fragments are shorter than 50 base pairs (bps) (Dabney 2013). This enrichment of orthologous sequences increases economy of sequencing and enables phylogenetic comparisons between individuals from samples of varying quality.

1.6 Aims and Structure of the Dissertation

The overall objective of my dissertation research was to develop capture hybridization techniques to characterize the genetic charcteristics and variation of Sunda and Philippine colugos from museum samples, calculate times of divergence, and hypothesize how colugo populations may have differentiated into their current population and geographic distibutions. This work represents the first substantial population based molecular characterization of Sunda and Philippine colugos, and the first genome-wide phylogenetic comparison to determine the proper placement of colugos within the eutherian phylogenic tree. In Chapter II I present our findings on

efficient capture hybridization-based recovery of DNA extracted from colugo museum specimens as published in *Genome Research* (Mason et al. 2011). Chapter III presents interordinal and intraordinal phylogenetic structure and proposal of colugo species groups to accurately describe the present genetic variation and is a manuscript submitted to the jounal *PNAS*. Chapter IV describes in greater detail the comparative biogeography of colugos, mouse deer, pangolins, and other Southeast Asian taxa and is in preperation for submission to the *Journal of Biogeography*. In this Chapter (IV) hypotheses are developed to explain how and when colugos may arrived at their present distributions via paleo-forest corridors. The final chapter (Chapter V) is a summary of the previous four chapters and furnishes a description of future scientific directions.

CHAPTER II

EFFICIENT CROSS-SPECIES CAPTURE HYBRIDIZATION AND NEXT-GENERATION SEQUENCING OF MITOCHONDRIAL GENOMES FROM NON-INVASIVELY SAMPLED MUSEUM SPECIMENS*.

2.1 Introduction

The advent of next-generation sequencing technologies (NGSTs) has transformed the way in which scientists approach a myriad of biological questions (Hawkins et. al. 2010). Even with NGSTs growing familiarity and broad range of applications such as ChIP-Seq, RNA-Seq, and genome wide association studies, NGSTs still have potential to influence the field of population genetics and phylogenetics with new methods to obtain genomic sequences of rare, difficult to sample, or extinct species (Millar et al. 2008). The ability to uncover the phylogenetic history of recently extinct species has rapidly improved due to the reduced cost and increased sequence capacity of NGSTs (Gilbert et al. 2007, 2008; Miller et al. 2008, 2009), however obstacles do remain. The difficulties with applying NGST's to phylogenetic problems do not lie with the sequencing technology itself, but with the preparative procedures for isolation and sequencing of large, orthologous DNA regions across multiple divergent species

---

* Reprinted with permission from Mason, V.C., Li, G., Helgen, K.M. & Murphy, W.J. (2011). Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Research* **21,** 1695-1704 by Cold Spring Harbor Laboratory Press under the Creative Commons License (Attribution-NonCommercial 3.0 Unported License).

(Summerer 2009). This problem is exacerbated for museum specimens where DNA quality varies greatly between samples and contamination levels are often high (Millar 2008). Generation of whole genome sequences for museum specimens, or even complete mitochondrial DNA (mtDNA) genome sequences, is not cost effective for most laboratories due to the large amount of sequencing required for adequate genome coverage of a single individual.

Capture hybridization methods are routinely utilized for genomic scale enrichments of modern target DNA from the same species (Summerer 2009, Mamanova et al. 2010), and also for recovery of DNA from museum or fossil specimens by largely removing contaminants from the final product (Krause et al. 2010). However, capture hybridization techniques have not been applied to assembling phylogenetic datasets across divergent sets of taxa (e.g. millions of years of genetic divergence), largely due to lack of appropriate probes and lack of exploration of hybridization conditions to allow for heterologous sequence capture. Enrichment for target sequences by PCR (which has been the standard for most previous museum DNA studies) requires closely related reference sequences and painstaking efforts to design many oligonucleotide primers to amplify very short regions of the DNA of interest. Capture hybridization and sequencing of targeted loci from museum specimens promises to be a more flexible, cost-effective, and efficient approach than other enrichment procedures for degraded samples. Here we describe the application of capture hybridization and selection techniques to recover mitochondrial DNA from thirteen Sunda colugo (*Galeopterus variegatus*) museum specimens of varying ages (47-170 years old) that represent major geographical

locations throughout the Southeast Asian mainland and archipelago (Table 2.1, Figure A2.1).

Colugos are arboreal mammals that are widely distributed throughout Southeast Asia, and have the most extensive gliding membrane (patagium) of any known mammal. This allows them to glide for very large distances, the longest recorded being 136m (Lim 2007). Colugos are rarely kept in captivity and are elusive in the wild (Lim 2007), factors that have obscured their evolutionary history for decades. Under the current taxonomy colugos comprise a unique mammalian order (Dermoptera) and are classified as two species: the Sunda colugo, *Galeopterus variegatus*, and the Philippine colugo, *Cynocephalus volans* (Wilson & Reeder 2005). However, recent mtDNA and nuclear DNA data provide compelling evidence that the geographically widespread

Sunda colugo in fact represents multiple species distributed throughout Southeast Asia (Janečka et al. 2008), and suggest further genetic sampling may identify many additional divergent populations and/or species. Because of the extreme difficulty obtaining fresh tissue or DNA samples from colugos we further explored this question utilizing collections of museum specimens and devised a comprehensive method for capture, selection, and recovery of divergent mtDNA fragments using NGST.

Table 2.1. USNM Sunda colugo specimens. Samples taken were dried adherent tissue.

I.D.	Tissue	USNM #	Date Collected	Location Sampled	Latitude	Longitude
1	Nasal cavity tissue	154600	25 May 1909	West Java, Mt. Salak	6°45'	106°41'E
2	Rib cartilage/tissue	155363	6 Apr 1909	East Java	ca 8° S	ca 113° E
3	Rib cartilage/tissue	307553	28 Sep 1957	Malaysia, Mt. Brinchong	4.52° N	101.38° E
4	Brain tissue	311297	17 Jul 1958	Malaysia, Langkawi Isl.	6°19'48 N	99°43'43 E
5	Rib cartilage/tissue	197203	2 Jul 1913	Borneo, Labuan Klambu	1.23° N	118.73°E
6	Nasal cavity tissue	317119	23 Sep 1960	Borneo, Ranau	5°57'8 N	116°39'52 E
7	Nasal cavity tissue	356666	8 Feb 1963	Thailand, Amphoe Kapoe	10° N	9.5° E
8	Nasal cavity tissue	198051	12 Jan 1914	Borneo, Kari Orang	0.83° N	117.87° E
9	Brain tissue	104600	7 Jul 1900	Natuna Isl., Sirhassen Isl.	2°31'13 N	109°2'51 E
10	Brain tissue	115605	20 Aug 1902	Sumatra, Rhio Arch.	1°1'31 N	104°27'44 E
11	Skull tissue	121749	12 Feb 1903	Sumatra, Batu Islands	0°25'26 S	98°26'47 E
12	Brain tissue	143327	12 Mar 1906	Sumatra, Pulo Rupat	1°52'32 N	101°34'48 E
13	Brain tissue	003940	1838-1840	Singapore	1°21'19 N	103°59'16 E

2.2 Results

2.2.1 Amplified Adapter-Ligated Museum DNA

Extraction of DNA from tissues of thirteen colugo museum specimens (Table 2.1) yielded varying amounts and qualities of DNA. DNA was recovered from every specimen, including the oldest, which was collected ~170 years ago (Table B2.1). Due to variation in the initial quantity between samples, we measured the degree of degradation from the PCR-amplified adapter-ligated DNA (Fig. 2.1). The age of museum specimens showed little correlation with quality of DNA recovered (Fig. 2.1). Quality of DNA was measured based on the size and intensity of the amplified, adapter-ligated-DNA smear on an agarose gel. Several specimens that were nearly 100 years old had higher molecular weight DNA when compared to specimens collected more recently. For example, specimen USNM 121749 (11), collected in 1903, showed comparably high quality DNA, while specimen USNM 356666 (7), collected in 1963, yielded among the poorest DNA qualities. Furthermore, DNA quality did not appear to be influenced by the source tissue, and was highly variable within and between tissue types, even when considering collection age. Thus, the quality of DNA recovered from each specimen seems to be largely influenced by how the specimen was handled and stored during and after collection rather than simply its age or tissue source.

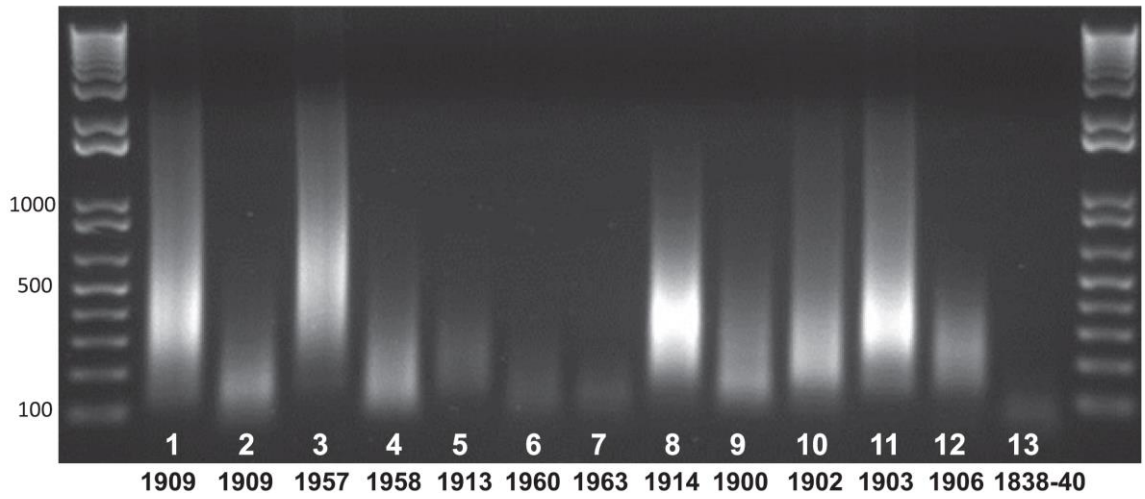


Figure 2.1. PCR amplified, adapter-ligated museum DNA extracts (5 µl) for 13 colugo specimens, resolved on a 1% agarose gel, with year of collection indicated below. Note that DNA quality does not always correlate with the age of each specimen.

2.2.2 Hybrid DNA Capture

The distribution and banding pattern of selected mtDNA fragments after two rounds of hybrid capture and amplification (2˚-selected, amplified museum products) is shown in Figure 2.2. There is a strong correlation between size and uniform distribution of the capture DNA smear with the percent genome coverage eventually obtained by Illumina DNA sequencing (Fig. 2.2 and Table 2.2). Specimens yielding uniform smears yielded the highest overall genome coverage percentages because they have a more even distribution and concentration of products. Specimens that yielded more banded 2˚ selected amplified products yielded more biased mtDNA genome sequence coverage, with a larger number of gaps.

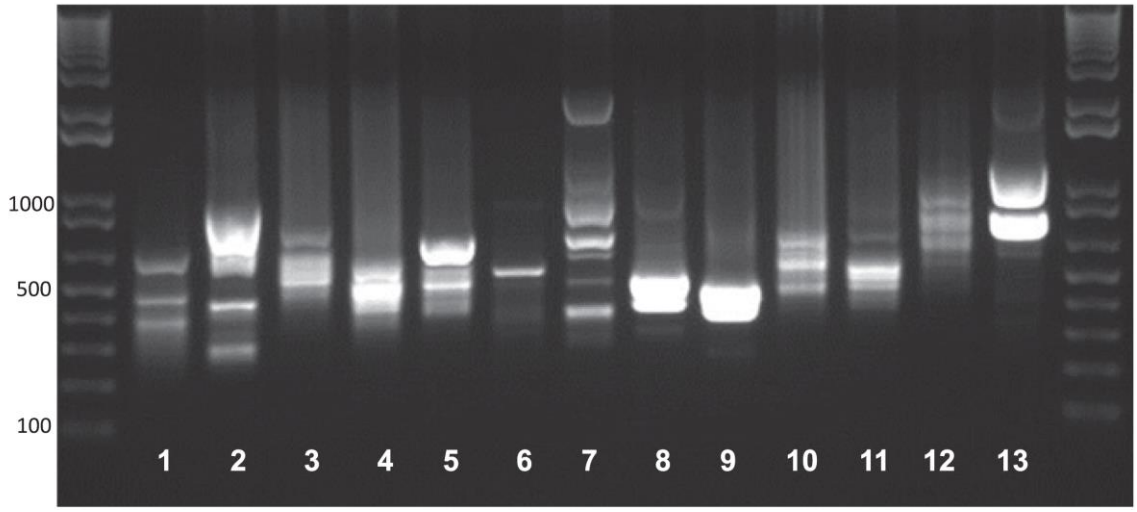


Figure 2.2. 2° selected and amplified mtDNA from the 13 museum specimens (5 µl) resolved on a 1% agarose gel.

Table 2.2. Results from the Illumina NGS of 12 indexed samples pooled together in 1 lane of a GAII flow cell and additional reads from a second multiplexing run that included USNM 104600 and USNM 003940. Note: The total # of reads indicates the number of reads after quality filtering and removing sequences under 30bps, but before removal of human contamination. Genome % coverage refers to the final coverage of each mtDNA genome in the alignment, with respect to the reference genome, and only includes sites with ≥5X coverage.

Specimen ID Number	USNM Number	Mapped to Reference:	Total Number of Reads	Number of Reads Mapped	Selection Efficiency[1]	Percent Human mtDNA	Genome Percent Coverage
1	154600	AF460846	119,351	107,063	92.60	0.008	89.96
2	155363	AJ428849	46,964	37,864	87.09	1.000	47.49
3	307553	AJ428849	91,182	71,543	83.65	0.280	90.39
4	311297	AJ428849	59,645	55,559	95.41	0.410	77.89
5	197203	AF460846	4,279	3,349	84.20	0.980	28.65
6	317119	AF460846	870,126	540,278	84.44	0.070	60.18
7	356666	AJ428849	403,791	154,803	41.11	0.020	22.73
8	198051	AJ428849	216,657	182,026	90.03	0.003	27.91
9	104600	AJ428849	654,203	444,647	67.97	0.290	13.68
10	115605	AJ428849	213,585	186,288	92.44	0.420	94.38
11	121749	AJ428849	103,328	90,085	92.54	1.960	70.58
12	143327	AJ428849	380,701	315,590	86.89	0.780	89.08
13	003940	AJ428849	580,924	9,417	1.63	0.047	18.86
		Total:	3,784,736	2,198,512			
		Average:	289,071	169,116	76.92%	0.48%	56.29%

[1]The total number of mtDNA reads that mapped to the reference genome + non-mapped reads with best hits to colugo mtDNA in GenBank. See Supplemental Information for further information on selection efficiency and BLAST results.

2.2.3 Illumina Sequencing

To make a preliminary evaluation of the efficiency of our selection procedure, we incorporated Illumina index sequences into the 2° selected-amplified museum products and created sequence libraries for two specimens: USNM 143327 (12) and USNM 317119 (6). These libraries were cloned, and 96 colonies sequenced per library using standard Sanger capillary techniques, and aligned to a published colugo reference

mtDNA genome (Fig. A2.2). This pilot study showed that between 70-90% of fragments recovered were of colugo mitochondrial origin, confirming the high selection efficiency of the capture procedure.

In light of these results we indexed the remaining selection libraries and pooled 12 of the 13 specimens in a single lane of an Illumina GAII flowcell. A single-end read, 84 cycle run returned an average of 24.4 million bps of sequence per individual (Table 2.2). After quality filtering and removing sequences under 30 bps, on average 76.92% of captured sequences showed strong similarity to the reference colugo mtDNA molecules (referred to hereafter as selection efficiency), 0.48% were of human mtDNA origin, and the remaining 22.60% represent other exogenous DNA (i.e. bacterial), nuclear DNA, or colugo sequence that was too divergent to map to the reference genome (Table B2.2). When the oldest specimen, 13, is excluded the average selection efficiency increases to 82.96%. Overall selection efficiency appeared to be influenced by starting DNA quality, however this was only pronounced in highly degraded samples, <150-200bp (Figs. 2.3A & B), as well as samples over 110 years old. By contrast, age of samples was not a good predictor of genome coverage (Fig. 2.3D).

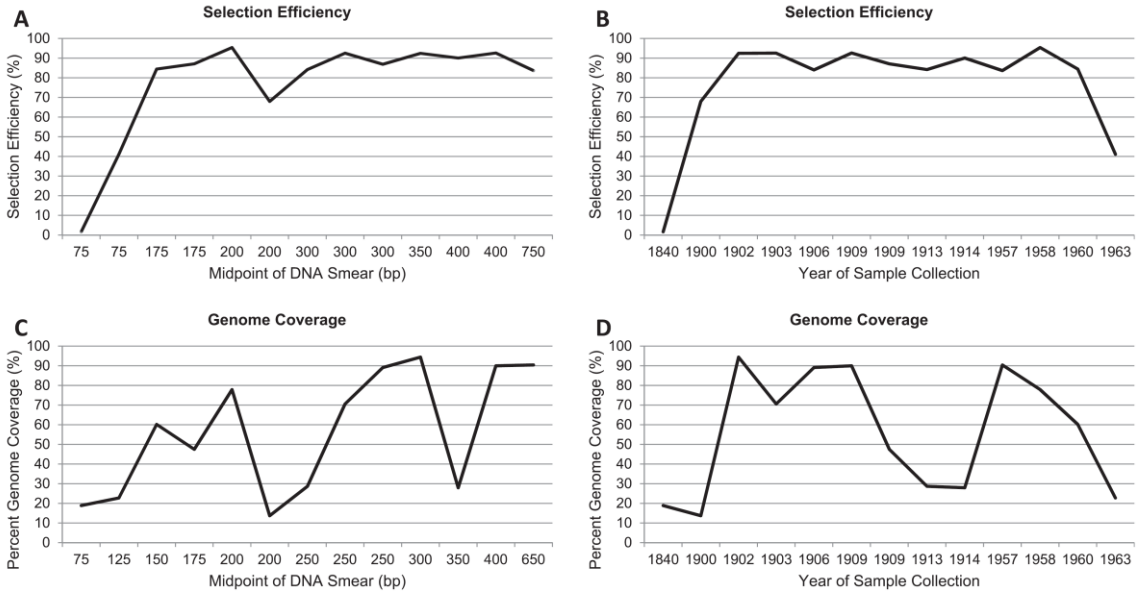


Figure 2.3. Plots of selection efficiency and mtDNA genome coverage relative to the original sample DNA quality (average size in bp of DNA fragments on 1% agarose gel [Fig. 2.1]), and age of sample.

Hybrid capture yielded an average depth/site of ~979x coverage (Table B2.3); this extreme depth allowed for accurate calling of 99.99% of bases used in our analysis. Captured sequence fragments were not evenly distributed across the genome in every individual: some mitochondrial genomes were nearly complete with >90% genome coverage, while other genomes had as little as 20% genome coverage (Table 2.2 and Figure 2.4). Across samples, conserved gene regions, such as the 12S and 16S rRNA genes and the conserved portion of the control region, possess higher depth of coverage than other areas of the mtDNA genome. On an individual basis, there was no obvious correlation between gaps in genome coverage and regions of low overall conservation across the genome, as assessed by the divergence plot between the bornean and javan

reference mtDNA genomes (Figure 2.4c). This would suggest that probe bias was minimal.

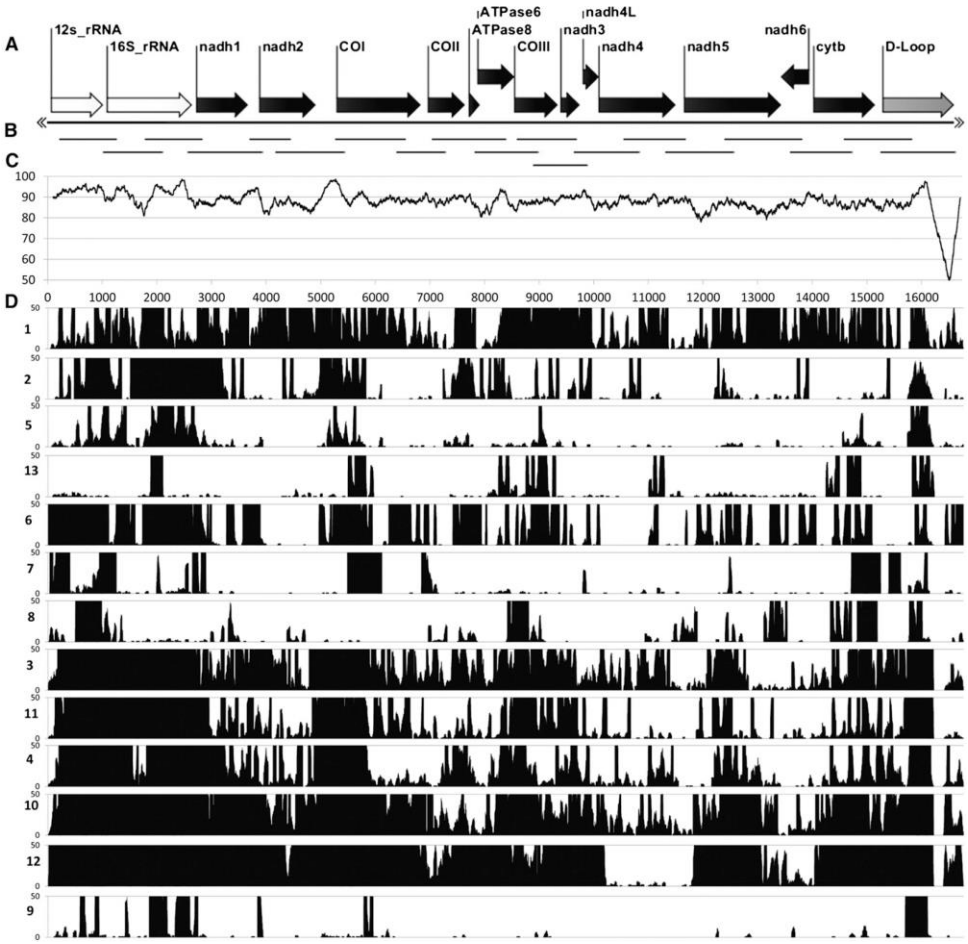


Figure 2.4. Distribution and depth of coverage of captured mtDNA fragments from each Sunda colugo museum specimen, displayed relative to the reference colugo mtDNA genome. (*A*) Reference colugo genome, AJ428849, depicted horizontally with gene annotations displayed except tRNA genes. (*B*) Distribution of mtDNA probe fragments produced by PCR. (*C*) Plot of DNA sequence identity between two full-length mitochondrial reference genomes (AJ428849, Borneo and AF460846, West Java) calculated in overlapping, 200-bp sliding windows. (*D*) Histogram showing distribution of captured mtDNA sequence fragments for the 13 museum specimens (labeled on left side by id number in Table 2.1), relative to the reference genome. Only coverage from 0-50X is shown for clarity, although most individuals have substantially higher coverage variation across the genome (Fig. A2.6).

2.2.4 Assessment of DNA Damage and Numts

Both ancient and historical samples are known to contain nucleotides that are damaged, the most common form being due to deamination of cytosine to uracil, which will lead to an excess of C-to-T and G-to-A substitutions when compared to a modern reference genome (Briggs et al. 2009, Millar et al. 2008, Krause et al. 2010). Our assessment of complementary C-to-T and G-to-A transitions versus T-to-C and A-to-G transitions from each individual sequence compared to a reference sequence, revealed no significant bias (α=0.05) in favor of transitions typical of chemical damage (Table B2.4B). This indicates that chemical damage has had little influence on our consensus sequences.

Numt's, or mtDNA that has been transposed into the nuclear genome (Triant and Dewoody 2007, Hazkani-Covo, Zeller and Martin 2010), can potentially be isolated by hybrid capture when they are similar to the probe sequence itself. The results of our initial Sanger sequencing identified 3 putative numts out of 183 high-quality Sanger reads, based on the presence of stop codons, immediately flanking nuclear sequence, or repetitive elements in the same sequence read. Though this suggests numts likely represent a very small fraction of captured sequences, we evaluated the 13 putative protein coding regions of the colugo mitochondrial genome in all final consensus sequences for stop codons or indels that might indicate capture of predominantly numt rather than cymt (cytoplasmic, or "true" mtDNA) sequences. We identified four nonsense mutations in four different individuals, and removed these sequence fragments as putative numts (Table B2.5). Not all numts are characterized by stop codons or indels.

Specifically, the recent transposition of numts into the nuclear genome could avoid detection by lacking indels or nonsense mutations (Hazkani-Covo et al. 2010). Therefore, we examined SNP frequencies and distributions within the read profiles of each individual, reasoning that high frequency SNPs that are clustered in specific regions of the mtDNA genome may be evidence of capture of recent numts (Fig. A2.5). Alternatively, SNPs that are shared in similar regions across individuals may indicate capture of ancestral numts. Although we did find evidence for high frequency SNPs across most individuals, the average frequency of these sites/individual was 0.42% (Table B2.6). We did not observe any significant stretches of SNPs of similar frequency that were clearly identified as numts. Furthermore, because of the multiple rounds of PCR amplification during the selection procedure it is difficult to exclude any SNP as a PCR-induced error incorporated during early rounds of the amplification process.

2.2.5 Phylogeny

The genetic divergence between the published reference genomes (AJ428849 and AF460846) and the final mtDNA consensus sequences of the study specimens (and GVA5 which served as our probe) were substantial, averaging ~9.0% for both AJ428849 comparisons (range 5.0%-13.7%) and AF460846 comparisons (range: 0.3%-13.2%) (Table B2.7). Notably, populations from Borneo, the Natuna Islands, and East Java were more divergent from the Bornean AJ428849 reference mtDNA genome than either the West Javan or Peninsular Malaysian populations (GVA1-6), which were considered divergent enough to warrant species level distinction (Janečka et al. 2008). Average within-island genetic distances were also very large: 8.1% between Bornean populations,

3.7% between Javan populations, 1.4% between Northeastern Sumatran islands, and 6.9% between both Peninsular Malaysia and Thailand populations (Table B2.7).

We constructed maximum likelihood (ML) trees for the complete mtDNA alignment of the 13 museum mtDNA sequences, two reference colugo mtDNA genomes (AJ428849 and AF460846), and six published partial mtDNA sequences from individuals from Peninsular Malaysia (GVA 1-3) and West Java (GVA 4-6) (Janečka et al. 2008). Enforcing different read depth thresholds for inclusion of a site in the alignment (e.g. read depth 5X-25X) had little effect on phylogenetic stability; all trees showed consistent clustering of the same major geographic lineages (Fig. A2.3). In light of these data we performed subsequent analyses on our minimum 5X depth dataset. The ML tree based on all sites (minus ambiguous regions in the alignment of hypervariable regions of the rRNA genes and control region) is shown in Figure 2.5. To minimize any effect of missing data on phylogenetic accuracy, we performed an analysis of different alignments where we varied the threshold for the number of individuals in the alignment that had sequence present at a given site (Fig. A2.4). Altering this parameter from 30-90% had little effect on well-supported nodes in Figure 2.5. Specimens from Peninsular Malaysia, Thailand, and several NE Sumatran islands formed a well-supported clade. This clade also consistently grouped with the Bornean reference genome (AJ428849) when the Natuna Island sequence (9) was excluded from analyses. The remaining Bornean populations formed a divergent clade that also includes a separate Thailand/Peninsular Malaysia group. A third divergent clade includes the West Java populations. Two other populations, East Java and the Natuna Islands, were not

consistently positioned within the phylogeny of colugos, possibly due to larger amounts of missing data and/or long branch attraction, and tended to lower bootstrap support for major geographic clusters (Fig. A2.4). Indeed, analyses that removed the Natuna specimen (9) showed increased bootstrap support for each of the major clusters shown in Figure A2.4. An analysis that only included sites present in the Natuna specimen (~2,250 aligned bp), and only individuals where >50% of these sites were present, revealed the Natuna colugo (9) to be a deeply divergent lineage with no close affinity to remaining colugo populations (Figure A2.4E). In summary, our results identify divergent phylogenetic groups of individuals from geographically distinct populations, confirming previous observations (Janečka et al. 2008).

2.2.6 Sequence Divergence and Capture Efficiency

Specimen 1 shows approximately 99% sequence identity to the full-length mtDNA genome of specimen GVA4, a mtDNA sequence obtained from preliminary assembly of the *Galeopterus* genome, as well as specimen GVA5, from which our biotin-labeled mtDNA probe was amplified. All three individuals represent populations within close proximity in West Java. Therefore specimen 1 serves as a good reference for estimating the maximum selection efficiency and genome coverage that might be obtained with our probe and hybridization procedure (assuming the DNA quality of this specimen is typical for what one might obtain from other museum specimens, Figure 2.1). Selection efficiency for individual 1 was ~93%, and genome coverage ~90%. Although these values are among the highest obtained for all of the museum specimens

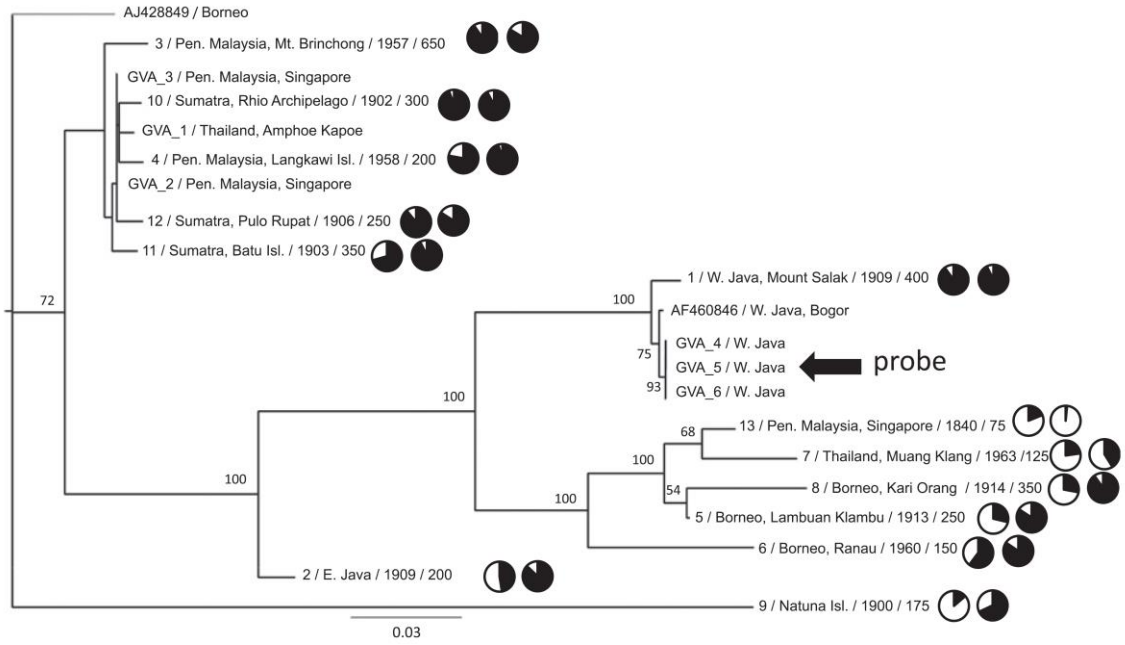


Figure 2.5. Maximum likelihood phylogenetic tree of Sunda colugo mtDNA sequences, and the effect of phylogenetic divergence on capture efficiency and genome coverage. Specimens are labeled sample numbers shown in Table 1, Texas A&M sample number (prefix "GVA"), or their GenBank accession number, followed by the geographic location of specimen collection. Additional sample information following the geographic location includes (left to right) date of collection, average starting DNA fragment size distribution (in bp), a pie chart showing % genome coverage, and a pie chart showing selection efficiency. Bootstrap values displayed at each node are based on 1,000 replicates. The tree shown is based on those sites where 50% of the individuals possess a base (14,008 bp, excluding hypervariable regions, see Methods) at a minimum depth of 5X. The overall relationships were supported in other analyses that minimized missing data and maximized data overlap across individuals (Fig. A2.4). The tree is displayed with midpoint rooting. The specimen (GVA4) from whom the mtDNA probe was derived is indicated with an arrow.

we analyzed, we obtained comparable selection efficiencies and genome coverage for individuals that were phylogenetically divergent from the probe sequence (Table 2.2, Figure 2.5). Specifically, hybrid capture from Sumatran and Peninsular Malaysian specimens (4, 10 & 11) achieved 92-95% selection efficiency, and similar levels of genome coverage (~90%) relative to specimen 1. Therefore our method can efficiently capture mtDNA sequences with as great as 10-13% sequence divergence from the probe sequence (Table B2.8), with minimal levels of mtDNA genome coverage bias due to probe divergence.

2.3 Discussion

Here we were able to demonstrate efficient cross-species capture of orthologous mitochondrial DNA sequence fragments, and in many cases nearly complete mtDNA genomes, from the degraded DNA of museum specimens collected >100 years ago. Below we discuss various experimental considerations for further improvement of capture hybridization across heterogeneous target loci and divergent species with unknown phylogenetic affinity, and examine the contribution of such data to addressing the phylogeny, taxonomy, and biogeography of a poorly known group of mammals, Sundaic colugos.

2.3.1 Cross-Species Probe Design

Probe construction is inherently important for successful recovery of target DNA, particularly when attempting cross-species hybridization with probes containing sequences with different levels of evolutionary conservation. While high sequence homology between the probe and target DNA allows for efficient and unbiased recovery

of target fragments among closely related populations of the same species (Mamanova et al. 2010), the greatest potential for capture hybridization lies with extinct species of uncertain phylogenetic affinity, or when there are no modern specimens available from an isolated, and possibly divergent, geographic population/species. However, as sequence divergence increases between the probe and the target individual, the potential to hybridize to exogenous DNA (e.g. numts, human contaminants) also increases. However, our hybridization conditions seem to be adequately relaxed to retrieve divergent (10-13%) mtDNA fragments, while maintaining high selection efficiency.

In the present study, we generated a probe from a single individual, although a pooled probe from multiple species or individuals from various locales might be more effective. This latter approach would lower annealing specificity but increase the probe's annealing potential (equivalent to a degenerate probe), which would be advantageous for probing taxa of uncertain phylogenetic affinity. By gradually relaxing the hybridization conditions, the touchdown approach (analogous to touchdown PCR) employed here provides more stringent and accurate hybridization conditions for conserved orthologous fragments of DNA to anneal prior to subjecting the probe to less specific annealing conditions. Occupation of probe fragments under stringent hybridization conditions removes or reduces the possibility for later mis-pairing during less accurate hybridization to paralogous DNA sequences, such as numts or human contamination. The touchdown hybridization approach appears to be extremely efficient as our levels of human and numt capture were considerably reduced relative to colugo mtDNA, while

allowing for capture (albeit unevenly) of sequences across the majority of the mtDNA genome from most specimens.

The objective of capture hybridization experiments is to obtain equal concentrations of every target base pair for equal sequencing depth coverage; hence another facet of cross-species probe production is controlling for varying levels of genome evolution across target regions. Though the high sequence coverage made it possible to recover orthologous fragments across the mtDNA genome of each specimen, as expected there was clearly a significant bias of sequence depth towards more conserved regions of the mtDNA genome (e.g. the 12S and 16S rRNA genes). This bias is particularly notable for the Natuna Islands specimen, which appears to be the most divergent Sunda colugo we sampled, and explains the more limited recovery of mtDNA fragments in more rapidly evolving parts of this specimen's mtDNA genome. To address the high coverage bias, future attempts might use multiple hybridization experiments, one probing for more conserved regions and the other for more divergent sections (based on pairwise sequence divergence), followed by equimolar pooling during NGS library production. In principle this strategy would obviate the need for touchdown hybridization, enabling different probe sets to hybridize longer at either stringent or relaxed conditions, thus providing greater opportunity to hybridize under optimal conditions. Alternatively, one could adjust the proportions of different probes/oligos in the pool, with more conserved regions represented in lower concentrations and less conserved regions in higher molar concentrations, in proportion to genetic divergence observed across related groups of taxa. This would allow more accurate hybridization

conditions for corresponding levels of divergence and more standardized representation of each base pair in the probe. Clearly, many variables need to be considered before attempting hybridization, depending on the extent of divergence anticipated between the probe and target individuals, allowing flexibility and customization in the hybridization capture experiment for successful recovery of a majority of the target DNA fragments.

2.3.2 Exogenous DNA

The initial quality of DNA derived from each museum specimen was evaluated based on the smear of DNA produced following amplification with adapter-ligated primers, and provided an estimate of the extent of degradation, and the potential for efficient hybridization. This initial DNA distribution could largely represent bacterial, or even human, contamination depending on how the specimen was handled and stored. Indeed, published reports of Next generation sequencing data from total DNA extracts of ancient mammalian hair or bone specimens indicate a substantial proportion of reads may correspond to exogenous DNA (Miller et al. 2008, Briggs et al. 2009). However, capture hybridization has the benefit of enriching for only target DNA, hence requiring fewer sequencing reads to obtain sufficient depth of coverage for accurate base-calling. Overall, our data showed very low levels of human as well as exogenous (bacterial) contamination following selection, though this did vary across specimens. Nonetheless, the levels of target colugo DNA sequenced were an order of magnitude greater than human contaminating DNA, and we achieved more than sufficient depth of coverage (>1000-fold in most cases) such that additional pooling of samples/flowcell lane is feasible. In this study, only two museum specimens (7 and 13) yielded less than 50%

captured colugo mtDNA, while the remainder shows a majority of reads to be of colugo origin. Even in those cases where colugo DNA did not represent the majority of captured reads, we could easily distinguish true colugo mitochondrial DNA through comparison of *de novo* contig assemblies with BLAST (Altschul et al. 1990) and alignment to the reference mtDNA genome sequences.

2.3.3 Colugo Phylogenetics and Southeast Asian Biogeography

Under current taxonomy, the Sunda colugo individuals sampled here from geographically widespread populations are classified as one species, *Galeopterus variegatus*, A previous genetic study that compared mtDNA and nuclear DNA fragments from specimens from Peninsular Malaysia, Borneo and West Java revealed a high degree of genetic divergence across colugo populations that exceeded levels observed for other pairs of well-established mammalian sister-species (Janečka et al. 2008). Our expanded analysis of genetic divergence between new and published mitochondrial sequences further indicates very large genetic divergence between specimens from geographically widespread localities such as Borneo, Peninsular Malaysia, Thailand, East and West Java, and the Natuna Islands. Although molecular divergence dates were not estimated on the current dataset (due to the absence of internal calibrations, and lack of a full length mtDNA genome from the closest outgroup, *Cynocephalus volans*), a previous study revealed potential species level distinctions between populations from West Java (GVA4-6), Borneo, and the Malay Peninsula (GVA1-3), with estimated divergence times between these populations as great as 5 million years (Janečka et al. 2008). Our results indicate similar or greater amounts of genetic divergence within and between

island populations (e.g. Borneo, Java, Natuna Islands), compared to values observed between the mainland and West Javan populations (Janečka et al. 2008) (See Figure 2.5; Table B2.6).

Fluctuating sea levels during the Pliocene and Pleistocene, along with other dynamic environmental changes, produced many isolating mechanisms that could promote speciation throughout the Southeast Asian archipelago (Harrison et al. 2006). Among these changes were the creation of many major river systems as well as expansions of dry savannah habitat. Colugos are obligately arboreal mammals and cannot survive in a savanna like ecosystem, nor can they traverse large rivers safely (Lim 2007). These geographic barriers likely generated at least four important refugia for arboreal mammals within the present-day landscape on the Southeast Asian continent (the "Sunda Shelf"): (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 probably extended from southern and eastern Borneo south to eastern Java, effectively isolating Western Java, which may account for the large observed genetic divergence between colugos from both West and East Java (Figure 2.5). Borneo was similarly 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 ranges running from the center to the northeastern tip of the island, and contains many major river systems dissecting the southern portion of the island in particular. The abundance of geographic isolating mechanisms would have provided many opportunities for population subdivision and speciation, and

supports the unexpected phylogenetic distribution and deep divergence of the Bornean specimens examined here. Further sampling of individuals from throughout Borneo, as well as the entire Sunda Shelf, will allow for more precise delimitation of taxonomic boundaries and allow for better elucidation of biogeographic scenarios, and the role different isolating mechanisms have played in the divergence and radiation of colugos.

The elusiveness of colugos, their absence in zoos, and their very broad geographical distribution make them an extremely difficult group of species to obtain detailed population-wide sampling without utilizing museum material. This is also the case for numerous poorly known, threatened, and endangered species throughout the world, particularly those in the tropical rainforests of Southeast Asia. Our results illustrate the extreme power of harnessing untapped genetic data within archived museum specimens of unknown genetic divergence by cross-species capture hybridization, coupled with NGS, to make genetic inferences which otherwise would be difficult or logistically improbable. It is likely that further genetic sampling of colugo specimens from throughout the Southeast Asian mainland, within southeastern Sumatra, southern and western Borneo, eastern Java, Natuna Islands, and the recently described population from Laos (Ruggeri and Etterson 1998) may provide additional evidence for deeply divergent colugo lineages that may warrant species level distinction. Broad application of this approach to other taxa will further enhance our ability to accurately estimate the true number of species on Earth, a necessary step towards preserving living biodiversity.

2.4 Methods

2.4.1 Museum DNA Extraction

Small pieces (~5 mg) of dried adherent soft tissue were sampled from crania, nasal cavities or cartilage of museum specimens deposited in the Division of Mammals in the Smithsonian Institution's National Museum of Natural History (abbreviation USNM). Museum specimens were digested overnight in 200 µl Lysis Buffer (Qiagen) with 100 µg proteinase K, followed by protein precipitation and isopropanol precipitation of genomic DNA following the general guidelines of the Gentra/Puregene DNA isolation protocol (Qiagen). The final elution volume for the DNA was 40 µl. DNA extractions and all-pre selection procedures were performed in a dedicated pre-PCR lab space for historic specimens, removed from the PCR/molecular biology laboratory.

2.4.2 Blunt Ending of Museum-DNA Extracts

DNA extracts were blunt-ended for adapter ligation using a 1 x 20 µl master mix of 9.5 µl H_2O, 8 µl of 5X reaction buffer for T4 DNA Polymerase (lot no. 0030577), 2.5U T4 DNA polymerase (Fermentas, lot no. 00032793) in the presence of 0.4mM dNTPs. 20 µl of the master mix was added to 20 µl of museum DNA extracts and incubated at 70˚C for 10 minutes. Following incubatiuon, extracts were purified with CentriSpin20 columns (Princeton Separations) returning ~32µL of product from a 40µL elution, stored on ice, and immediately followed by ligation of adapters.

2.4.3 Adapter Ligation and Amplification of Museum DNA

Ligation of double-stranded adapters (generated by self-ligation of two oligos ORM-28 and ORM-29; Peterson 1998) to 32 µl of blunt ended museum DNA extracts were performed in a 150 µl ligation volume, using 2.5U T4 DNA ligase (Fermentas) and 1X T4 DNA ligase buffer + ATP (Fermentas), incubating 19 hours at 16˚C. PCR amplification of the resulting adapter-ligated museum fragment libraries was performed using the ORM-28 primer (0.5 µM) in 5 x 50 µl reactions with 2.5U Platinum *Taq* DNA polymerase (Invitrogen), 1x PCR buffer, 1.5 mM $MgCl_2$, and 0.8mM dNTPs. The PCR profile included a 1 minute hot start at 95°C, followed by 20-30 cycles (depending on starting concentration of DNA) of denaturing for 15 seconds at 94°C, annealing for 20 seconds at 58°C, and extension for 1 minute at 72°C, followed by a final 5 minute extension at 72°C. PCR reactions from each individual were pooled, purified in a MicroconPCR device (Millipore) and resolved on 1% agarose gels to examine fragment size distribution.

2.4.4 Capture Probe Generation

A high-quality West Java colugo tissue specimen (GVA5) served as the source of DNA for generating our mtDNA probe. To increase the probability that the mtDNA amplicons were of mitochondrial origin, rather than a nuclear mtDNA pseudogene (numt), we used a mitochondrial enrichment procedure to generate template DNA (Jones et. al. 1988), modified to allow for small-scale extractions. 1.5 mg of liver 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 for 15 minutes in a 4˚C microcentrifuge to pellet the nuclear debris. Following transfer of the supernatant 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 supernatants were combined and spun at 13,400 rpm for 30 minutes at 4˚C to pellet the mitochondria, and the nuclear pellet was preserved at -80˚C. The mitochondrial pellet was resuspended in 200 µl of Cell lysis buffer (Qiagen), digested overnight at 56˚C with 100 µg of Proteinase K, and cooled on ice for 7 minutes. Sixty seven microliters of protein precipitation solution (Qiagen) was added, vortexed for 20 seconds, and spun for 5 minutes at 8000 rpm. The supernatant was transferred into a fresh tube to which 200 µl of 100% isopropanol was added, inverted 50 times, and spun for 15 minutes at 12000 rpm. The DNA pellet was washed with 200 µl of 70% ethanol, spun for 2 minutes at 12000 rpms, and allowed to air dry for 5-10 minutes. The DNA was eluted in 20 µl of Elution buffer (Qiagen). The enriched mtDNA extract served as a template for amplifying 19 ~1-1.4 kb, overlapping fragments (in triplicate) that span the colugo mtDNA genome (Fig. 2.3, Table B2.9). PCR cocktail in 3 x 25 µl reactions with designed forward primer (2 µM), designed reverse primer (2 µM), 0.5U Platinum *Taq* DNA polymerase (Invitrogen), 1x PCR buffer, 1.5 mM $MgCl_2$, and 0.8mM dNTPs. The PCR profile was: 2 minute hot start at 94°C, followed by 40 cycles of denaturing for 15 seconds at 94°C, annealing for 30 seconds at 60-50°C where the annealing temperature drops during the first 10 cycles by 2˚C every other cycle, and extension for 1 minute at 72°C, followed by a final 2 minute extension 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. PCR products for each amplicon were pooled and purified individually using MicronPCR devices (Millipore), and quantified. Fragments were then pooled based on concentration and fragment length to obtain equal representation of each base pair in the mitochondrial genome. The pooled mtDNA probe was labeled by biotin-nick translation (Roche) or biotin-High Prime (Roche) random priming procedures, following manufacturer protocols.

2.4.5 Capture Hybridization and Selection

Capture hybridization follows a modified version of Del Mastro and Lovett's (1997) protocol, originally described for cDNA selection with a genomic probe, with minor changes. Approximately 500 ng-1µg of amplified, adapter ligated DNA was combined with 100 ng of biotin labeled mtDNA probe and added to an equal volume of 2X Hybridization Buffer (1.5mM NaCl, 40mM Sodium Phosphate Buffer [18.25 ml 1M NaH_2PO_4, 77.75 ml of 1M Na_2HPO_4], 10mM EDTA, 10x Denhardt's solution, and 0.2% SDS), not exceeding 15 µl. The sample was overlaid with 50 µl of mineral oil, denatured for 5 minutes at 99°C, and incubated for 50 hours at 65-60°C, reducing ~2°C every 24 hours. This "touchdown" approach was used to enhance retrieval of more divergent DNA sequences relative to the capture probe. Following hybridization, samples were added to 1 mg of Dynabeads M-280 Streptavidin beads (Dynal), which had been washed three times in 100µl TEN buffer (10mM Tris-HCl, pH 7.5, 1mM EDTA, 1M NaCl) utilizing a magnetic tube holder, and resuspended in a final volume of 100µl TEN. Two room temperature low stringency (1X SSC, 0.1% SDS) and three 65°C high stringency

(0.1X SSC, 0.1% SDS) washes were performed following DelMastro and Lovett (1997). The beads were eluted in 25µl 0.1N NaOH at room temperature for 20 minutes, with gentle vortexing every 5 minutes, neutralized with 25 µl of Tris-HCl, pH 7.5, and the total 50 µL was passed through a Centri-spin 20 column (Princeton Separations). The primary selected DNA was amplified in four replicate 50µl PCR reactions using 10µl DNA, 0.5 µM ORM-28 primer, 2.5U Invitrogen Platinum Taq in 1x PCR buffer, 1.5 mM MgCl2, and 0.08mM dNTPs, and the following PCR profile: 1 minute hot start at 95˚C, 35 cycles of 15 seconds at 94˚C, 20 seconds at 58˚C, 1 minute at 72˚C, followed by a 5 minute final extension at 72˚C. PCR products were pooled and purified with Montage-PCR filters (Millipore) and resolved on 1% agarose gels to examine the size-distribution of selected DNA fragments. A second round of DNA capture and selection was repeated using 1 microgram of 1˚ selected, amplified DNA as template, and 100 ng of the biotin-labeled capture probe, using the same procedure described for the primary selection and amplification. Final amplification used 5 µl of template DNA per reaction, and 30 rounds of PCR amplification.

2.4.6 Initial Evaluation of mtDNA Selected Libraries Using Sanger Sequencing

The indexed libraries for specimens 6 and 12 were cloned into the PCR-TOPO Blunt end vector (Invitrogen) and grown on LB+ampicillin plates. 96 colonies from each library were picked into 15 µl of sterile 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 sequenced (ABI Big Dye3.1) on an ABI-3730 capillary DNA sequencer (Agencourt

Biosciences). The resulting DNA sequences were edited for quality and vector+adapter-trimmed in Sequencher (Genecodes, Inc.). The DNA sequences were then assembled relative to a colugo reference mitochondrial DNA genome (AJ428849).

2.4.7 Next Generation Sequencing and Sequence Assembly

MtDNA selection products in the 200-600 bp range were gel excised and higher intensity bands outside the main fragment smear were excised, cloned, and sequenced separately with Sanger-based sequencing to reduce bias in sequencing coverage across fragments in the library. Following standard Illumina specifications, the main fragment smear was subsequently purified, and indexed by PCR with Illumina paired-end primers where 12 unique index sequences were incorporated into the paired-end adapters to distinguish sequences from different individuals. These twelve individuals were multiplexed in one lane of an Illumina GAII flowcell, and later sorted computationally (Table 2.2). Sequences for USNM 003940 (13), and additional reads for USNM 104600 (9) to increase representation from low coverage regions, were generated in a second lane of a separate run. 3.5 million 84-bp reads were generated from the first reaction, while the addition of the colugo reads in the second run increased the total to 4.3 million reads (Table 2.2).

Sequences were trimmed of ORM-28 adapter sequences by removing the first 22bps of each sequence. Sequences were trimmed by quality both in CLC Genomics Workbench with default parameters, and in EULER-SR (Chaisson and Pevzner, 2008). We further queried and masked all ORM-28, ORM-29, and Illumina adapter sequences in the sequence reads. Artifacts from the various ligation procedures were identified and

sequences containing them were removed. Sequences were then imported into CLC where all sequences < 30bps were removed.

To identify and remove human sequence contamination, the high quality reads were mapped to a human reference sequence (HM125971) where 100% of each read must match at 98% similarity to the reference sequence. The remaining unmapped reads were then mapped under global alignment parameters to two colugo reference sequences (AJ428849 and AF460846) where 100% of each read must match at 85% similarity to the respective reference sequence. Consensus sequences and depth information were derived from these alignments in CLC. Sequence depth parameters were enforced using a custom Perl script.

2.4.8 Open Reading Frame Analysis

Open reading frames (ORFs) of each individual were analyzed by translating coding domain sequences (CDS) regions and checking for premature stop codons, as well as comparing translated regions to all previously published CDS regions of a colugo mtDNA genome (AJ428849). Analyses were performed in CLC using the vertebrate mitochondrial genetic code.

2.4.9 Phylogenetic Analysis

Sequence alignments were performed in Sequencher (vers. 4.8, GeneCodes, Inc.) and adjusted by eye. Hypervariable regions of ambiguous alignment were excluded from further analysis. Maximum likelihood trees were generated with RAxML (vers. 7.0.3, Stamitakas et al. 2006), under a GTR+gamma model of sequence evolution. Bootstrap support metrics are based on 1,000 replicates. Pairwise genetic distances and other

sequence statistics were generated in MEGA 5.0 (Tamura et al. 2007), using maximum composite likelihood distances (gamma corrected). To investigate the effect of depth on phylogenetic robustness (i.e. potential errors in low-coverage regions might influence phylogenetic accuracy) we constructed ML trees from different alignments where inclusion of a site in the alignment required a specific read-depth: 5X, 10X, 15X, 25X (Figure A2.3). ML trees were also constructed from alignments where 30, 50, or 70 percent of individuals share a base at that base site (Figure A2.4).

2.5 Data Access

Raw sequence data have been submitted to the SRA database under StudyAccession#: SRP007459, and Sample Accession numbers SRS214574 and SRS214579-SRS214590. The sequence alignment has been deposited in TreeBase: http://purl.org/phylo/treebase/phylows/study/TB2:S11695

CHAPTER III

GENOMIC ANALYSIS REVEALS REMARKABLE HIDDEN BIODIVERSITY WITHIN COLUGOS AND THE SISTER GROUP TO PRIMATES

3.1 Introduction

As members of a strictly arboreal lineage of Southeast Asian gliding mammals, colugos (Order Dermoptera) have been known to science for centuries. However, the absence of captive individuals and a cryptic, nocturnal lifestyle have left basic questions surrounding their ecology and evolutionary history unanswered (Lim 2007, Lim et al. 2013, Stafford & Szalay 2000). At various times within the past century colugos have been allied to mammals as divergent as insectivores and bats, and have played a central role in discussions of primate ancestry (e.g., colugos are often erroneously referred to as 'flying lemurs') (Beard 1993, Jackson & Thorington 2012). Indeed, the phylogenetic position of colugos relative to other euarchontan orders remains highly controversial, with competing studies favoring an association of colugos with either primates or treeshrews (Schmitz et al. 2002, Janečka et al. 2007, Martin 2008, Meredith et al. 2011, O'Leary et al. 2013, Kriegs et al. 2007, Lin et al. 2014). Current taxonomy describes the order Dermoptera as one of the least speciose within all of Mammalia, consisting of just two species in monotypic genera: the Sunda colugo, *Galeopterus variegatus*, and the Philippine colugo *Cynocephalus volans* (Wilson & Reeder 2005). This low species richness is surprising because colugos are widely distributed throughout the Southeast Asian mainland and archipelago, a region of otherwise remarkable and rapidly

disappearing biodiversity (de Bruyn et al. 2014). Colugos also possess the most elaborate gliding membrane among living vertebrates, which inhibits terrestrial movement and dispersal outside of forests (Lim 2007, Lim et al. 2013). Population differentiation is supported by morphology (Stafford & Szalay 2000) and high mitochondrial divergence between some Sunda colugo populations (Janečka et al. 2008, Mason et al. 2011). These may reflect valid species isolated in allopatry, but remain unsubstantiated in the absence of broader geographic and genomic sampling. Thus, evolutionary questions surrounding dermopteran origins and taxonomic diversity remain unresolved, potentially influencing conservation strategies and the interpretation of early primate origins and evolution (Martin 2008, Moritz et al. 2013, Melin et al. 2016).

To produce the first detailed genetic insights into the poorly known history of this enigmatic mammalian order, we produced a draft genome assembly from a male Sunda colugo from West Java. We generated ~55× depth of coverage using Illumina sequence reads and produced an assembly (G_variegatus-3.0.2) that is 3.2-Gbp in length (see methods). This assembly is longer than most eutherian genomes, with a scaffold N50 of 245.2-Kbp and contig N50 of 20.7-Kbp. The assembly was annotated with the NCBI annotation pipeline and colugo RNAseq libraries (see methods), which identified 23,081 protein-coding genes. To test competing hypotheses concerning the relationship of colugos to other mammals, we performed comparative genomic analyses with a 2.5 Mbp one-to-one orthologous coding DNA sequence (CDS) alignment between colugo and seventeen other sequenced mammalian genomes (see methods, table B3.1). These alignments were augmented with reference assemblies from a male Philippine colugo

based on 14× Illumina sequencing coverage, and a pentailed treeshrew (Ptilocercus lowii) based on ~5× coverage, to mitigate long-branch attraction (LBA) effects (see methods).

3.2 Results and Discussion

Maximum likelihood and coalescent-based phylogenies constructed with nucleotide and amino acid versions of the genome-wide supermatrix consistently supported the Primatomorpha hypothesis (Beard 1993), confirming colugos as the sister-group of primates (Janečka et al. 2007, Meredith et al. 2011, Melin 2016) (Fig. 3.1, fig. A3.1). Because tree-building methods applied to deep, star-like radiations with shallow terminal lineages may be confounded by LBA artifacts (Lin 2014) we also searched the whole genome alignments for two independent types of phylogenetic character support that are not influenced in this way: 1) in-frame protein coding INDELs (insertion/deletions) and 2) near homoplasy-free retrotransposon insertions (see methods). We identified 20, 5, and 5 coding indels (chi-square p=4.5e-05) and 16, 1, and 0 retrotransposon insertions (KKSC test p=2.7e-07) supporting Primatomorpha, Sundatheria (colugos+treeshrews), and Primates+treeshrews respectively (Fig. 3.1, figs. A3.2-A3.4, table B3.2). These statistically robust reconstructions of Primatomorpha stand in stark contrast to the phenomic dataset of O'Leary et al. (O'Leary et al. 2013) who identified 69 morphological characters uniting colugos with treeshrews (Sundatheria). Our results imply that any morphological similarities uniting colugos with treeshrews (O'Leary 2013, Bloch et al. 2007) are due to convergent evolution or represent primitive euarchontan characters lost in the primate ancestor.

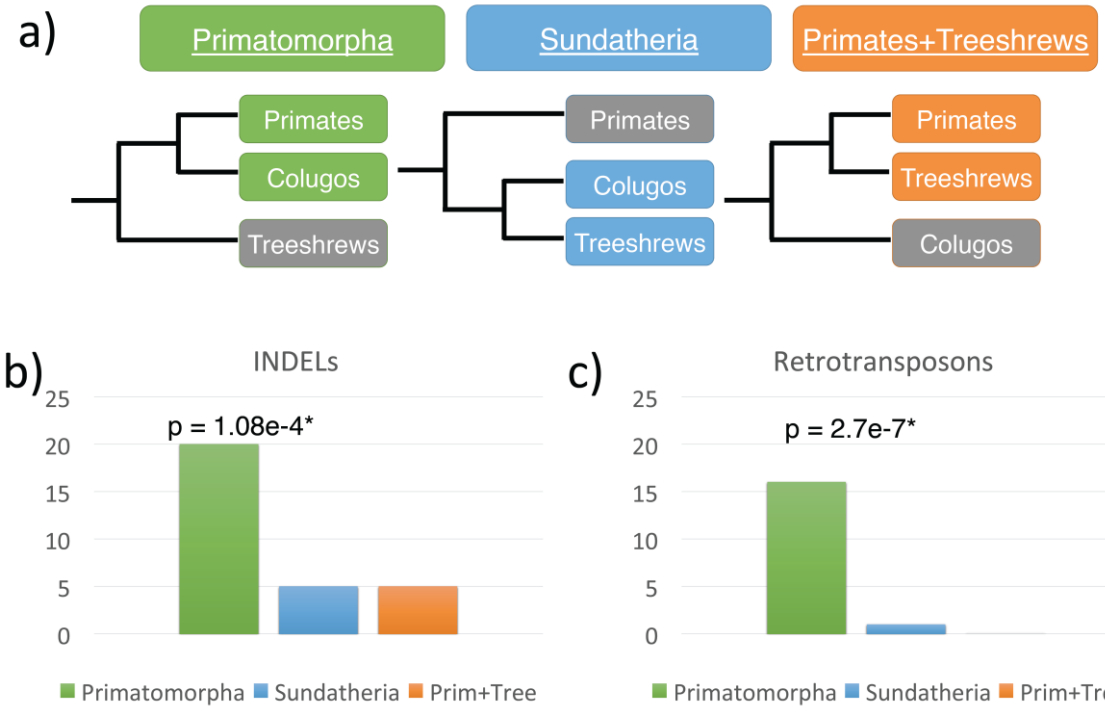


Figure 3.1. Phylogenetic placement of Dermoptera. A) Phylogenies depicting alternative hypotheses for dermopteran relationships relative to primates and treeshrews. B) Number of INDELs supporting each evolutionary relationship. C) Number of transposable elements supporting each evolutionary relationship.

We used a comparative genomics approach to explore the annotated gene sets of multiple euarchontans to investigate two major events in their early evolution: 1) lineage-specific genetic changes that plausibly support colugo adaptations and provide insight into their distinctive biology, and 2) lineage-specific genic changes that represent ancestral primate innovations. We annotated the colugo olfactory (OR) and vomeronasal (V1R) gene superfamilies (see methods), which encode odorant and pheromone receptors, and found that they were intermediate in size between treeshrews and primates

(Fig. 3.2a, Table B3.3). This finding supports a progressive loss of OR gene repertoires that began in the ancestral lineage of Primatomorpha. We also found evidence for an increased importance of vision and hearing in the colugo lineage based on significant enrichment for positively selected genes (PSGs) involved in these sensory modalities (padj =0.0097 and padj =0.004, respectively); i.e. genes that when mutated, are known to cause sensorial hearing loss and a variety of visual pathologies including macular degeneration (Fig. 3.2c, table C3.1 and table C3.2). The increased number of loss-of-function OR and V1R gene mutations in colugos is consistent with the view that selection for enhanced visual processing in a nocturnal, arboreal milieu corresponds with a relaxation of selection on olfaction (Wang 2010). The magnitude of this hypothesized tradeoff is greatest among mammals that experienced adaptive shifts from nocturnality to diurnality (Barton 1995), but here we show prevalence in a decidedly nocturnal lineage.

Positive selection was detected on similar vision-related genes on the ancestral primate branch, notably those in which mutations are implicated in night blindness and retinal degeneration (e.g., NXNL1, C8orf37). The ancestral primate branch also showed significant enrichment for PSGs that underlie brain function (padj =0.0001), including neurotransmitter genes implicated in behavioral disorders like schizophrenia and bipolar disorder, and neurodegenerative disease (padj =0.0004) (Table C3.3 and C3.4). The

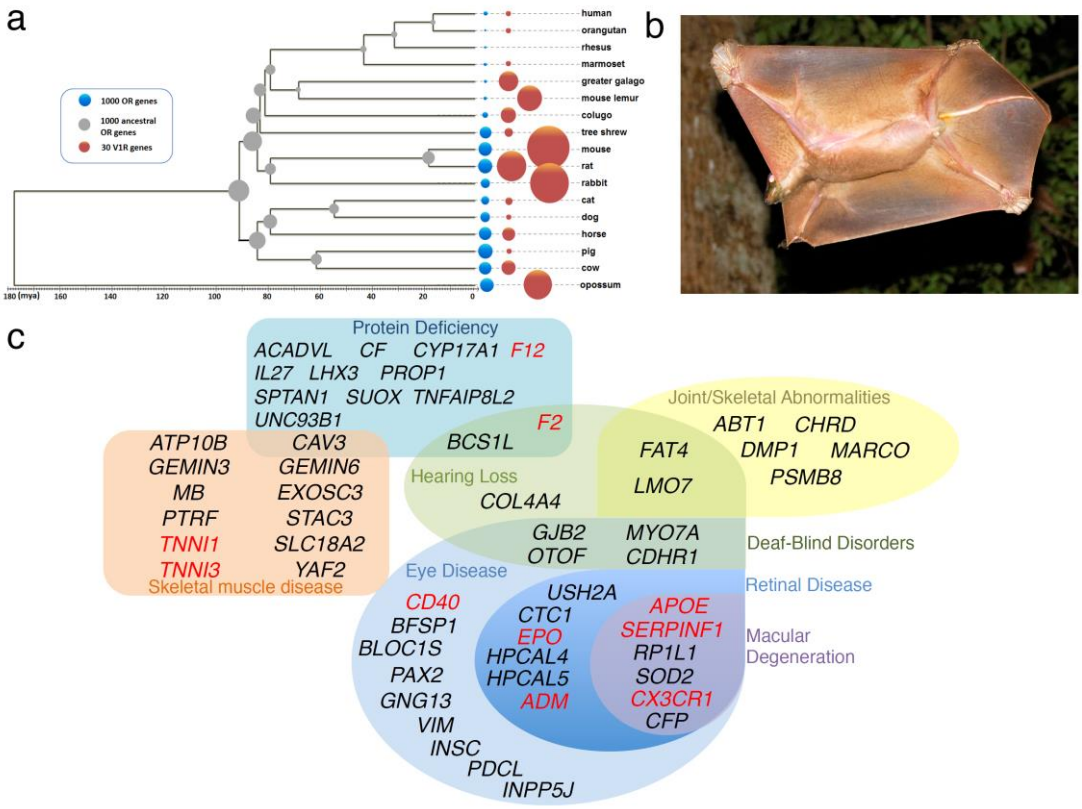


Figure 3.2. Functional gene evolution and positive selection in colugos and ancestral primates. A) Relative abundance of functional *V1R* (orange) and *OR* (blue) genes across sequenced mammals. The size of the circles is proportional the number of functional genes. B) Colugo gliding with patagium fully extended. C). Venn diagram showing relationship between categories of enriched gene categories of colugo positively selected genes.

latter category includes ATXN10 and SACS, two genes in which mutations are associated with autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS), a human genetic disorder characterized by early-onset spastic ataxia, nystagmus, distal muscle wasting, finger and foot deformities, and retinal hypermyelination (Storey 2014). It is highly plausible that positive selection on this suite of genes underpins the early morphological and behavioral evolution from ground-dwelling, scansorial ancestors to

early arboreal primates adept at grasping and climbing (Sussman et al. 2013). The colugo patagium is the most extensive gliding membrane of any living vertebrate, and stretches to the extremes of the digits and the tail when fully extended, resembling a living kite (Fig. 3.2b). Enriched disease gene categories within the dermopteran PSG set include muscular atrophy (padj =0.0086) and protein deficiency (padj =0.0002), including genes involved in muscle contraction (e.g., SLC18A2, TNNI1, TNNI3) (table C3.2). Eight PSGs are also associated with joint/digital deformities in a variety of disorders (table C3.2). We speculate that adaptive changes in this suite of genes contribute to the gross anatomical transformations of the musculature and skeleton that evolved in the arboreal ancestors of these skilled gliders (Beard 1993).

Colugos are widely distributed throughout Sundaland, a region well known for species richness and complex biogeographic patterns due to fluctuations in temperature, sea level, and vegetation throughout the Neogene (Cannon et al. 2009, de Bruyn et al. 2014). While current Sundaic forest distributions are in a refugial state with high sea stands, sea levels have been more than 40m below current levels for ~92% of the past one million years (de Bruyn et al. 2014, Cannon et al. 2009). Such exposure of the Sunda shelf connected many islands with the mainland and with each other. It has been difficult to decipher the geographical extent of forested connections during low sea stands because geological, biotic, and climatic evidence remains inconclusive. Sundaic phylogeography is potentially informative in this regard, however most widely distributed species that have been studied are either volant (e.g. birds or bats) or highly vagile (e.g. carnivores), and many have diversified very recently within the region

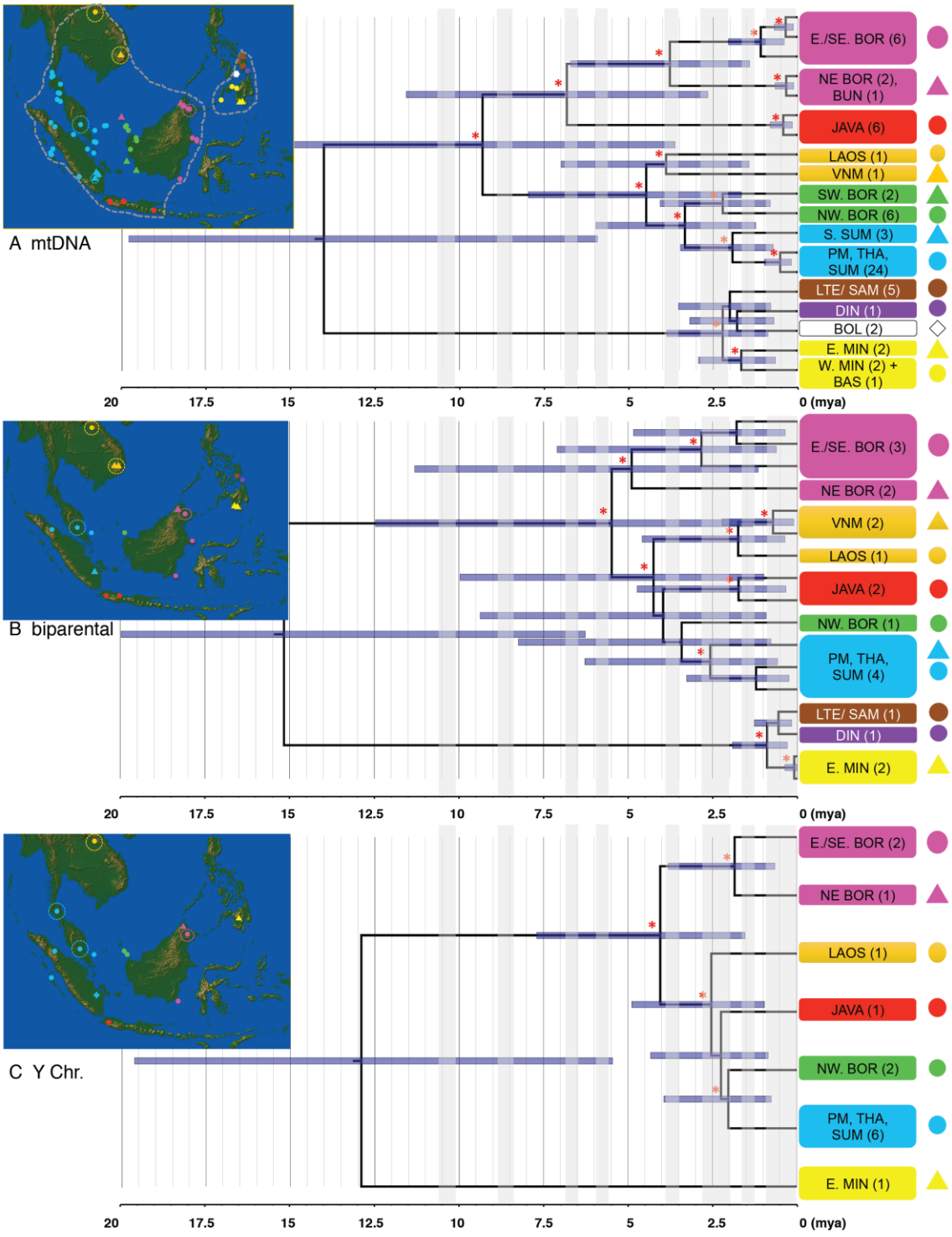
(Cannon et al. 2009, Bird et al. 2005). We hypothesized that colugos should track ancient Sundaic forest distributions due to their probable origin and widespread diversification within Sundaland, their preference for closed-canopy forests and reported isolation by rivers, disturbed forests, and savannahs (Lim 2007, Lim et al. 2013).

Given the scarcity of modern colugo samples with which to test this hypothesis, we exploited capture-based next generation sequencing technologies (Mason 2011) to retrieve orthologous DNA sequences (tables A3.4 and A3.5, table C3.5) from a broad sampling of museum specimens (table B2.8)(see methods) distributed across Sundaland and the southern Philippine islands of Greater Mindanao (Fig. 3.3A). We targeted ~140-Kbp of biparental and Y chromosome loci from 66 colugo museum specimens that were between 28-121 years old and yielded adequate DNA (Table C3.6, 19). We also obtained mitogenomic sequences from both off-target nuclear capture reads and direct low-coverage genome sequencing (table B3.6)(19). The colugo molecular time-trees were calibrated with the 95% confidence interval of our molecular estimate (avg.=10.9 mya) of divergence time between *Galeopterus* and *Cynocephalus* (figs. A3.5-A3.6, table B3.7)(see methods). Maximum likelihood-based maternal, paternal, and biparental phylogenies for both genera sort strongly by geography, showing major colugo lineages diversified in the Miocene or Pliocene (Fig. 3.3, figs. A3.7-A3.12). In addition, principal component analysis of 19 craniodental measurements (fig. A3.13-A3.14), as well as X-Chromosome SNP variation (fig. A3.15-A3.19), sorts Sunda colugos largely by geographic location (see methods). Notably, the large island of Borneo harbors multiple, deeply divergent colugo lineages, with eastern and western populations spanning the

oldest bifurcation within *Galeopterus* (Fig. 3.3, figs. A3.7-A3.12, tables B3.8-B3.9). This supports our prediction that ecological or topographic features such as mountains or major river systems presented substantial dispersal barriers to colugos, despite simulations that predict forested connections throughout Borneo up to the Last Glacial Maximum (LGM)(Cannon et al. 2009).

Ongoing debates argue for the presence/absence of a north-south savannah corridor separating Borneo from Peninsular Malaysia and Sumatra, which may have prevented dispersal of forest-dependent species while allowing dispersal of larger terrestrial mammals between Indochina and Java (Bird et al. 2005, Leonard et al. 2015). We observed complete sorting of colugo mitochondrial and nuclear haplotypes from Peninsular Malaysia/Sumatra and western Borneo (Fig. 3.3), suggesting an absence of Pleistocene genetic exchange despite evidence for a forested connection at the LGM (Cannon et al. 2009, Sheldon et al. 2015). We infer that forested dispersal corridors during late Pleistocene glacial maxima were fragmentary or rare, or that strong reproductive isolating barriers to gene flow had accumulated in allopatry throughout the Pliocene, limiting introgression and retaining geographic structure. In contrast, colugos from Thailand, Peninsular Malaysia, and Sumatra show less than 1.0% mitochondrial divergence from most of their satellite islands (table B3.10) supporting recent, geographically limited dispersal and colonization following repeated insular submergence during the late Pleistocene.

Figure 3.3. Colugo phylogeography based on museomics. Time trees based on major lineages within phylogenies representing A) maternal (mtDNA, 16.6 kb), B) biparental (autosome + chrX, 115.6 kb), and C) paternal (chrY, 24.3 kb) evolutionary histories (figs. A3.7-A3.12). Nodes with 100% ML bootstrap support are denoted with red asterisks. Maps depict sample collection locations for each tree with corresponding colored symbol. Boxes indicate highly supported monophyletic clusters or divergent independent lineages representing putative species. Grey shading denotes times of low sea stands. Dashed lines indicate the known distribution of Sunda and Philippine colugos. Abbreviations: Vietnam=VNM, Thailand=THA, Peninsular Malaysia=PM, Sumatra=SUM, Borneo=BOR, Pulau Bunguran=BUN, Pulau Basilan=BAS, Mindanao=MIN, Bohol=BOL, Leyte=LTE, Samar=SAM, Pulau Dinagat=DIN.


A mtDNA
E./SE. BOR (6)
NE BOR (2), BUN (1)
JAVA (6)
LAOS (1)
VNM (1)
SW. BOR (2)
NW. BOR (6)
S. SUM (3)
PM, THA, SUM (24)
LTE/ SAM (5)
DIN (1)
BOL (2)
E. MIN (2)
W. MIN (2) + BAS (1)
20 17.5 15 12.5 10 7.5 5 2.5 0 (mya)
B biparental
E./SE. BOR (3)
NE BOR (2)
VNM (2)
LAOS (1)
JAVA (2)
NW. BOR (1)
PM, THA, SUM (4)
LTE/ SAM (1)
DIN (1)
E. MIN (2)
20 17.5 15 12.5 10 7.5 5 2.5 0 (mya)
C Y Chr.
E./SE. BOR (2)
NE BOR (1)
LAOS (1)
JAVA (1)
NW. BOR (2)
PM, THA, SUM (6)
E. MIN (1)
20 17.5 15 12.5 10 7.5 5 2.5 0 (mya)

Similar to the separation of colugos from Borneo versus Java and Peninsular Malaysia/Sumatra, Philippine colugo mitochondrial lineages are private to different islands and coalesce to the early Pleistocene (>1.5 mya) (Fig. 3.3a, fig. A3.7-A3.9). These dates mirror similar genetically structured patterns in several Philippine mammals, including tarsiers (Brown et al. 2014) and Apomys rodents (Steppan et al. 2003), suggesting that current deepwater channels formed effective barriers to inter-island dispersal of arboreal lineages for much of the early Pleistocene. Nuclear gene loci also support monophyly of sampled islands (Fig. 3.3b, fig A3.10), but divergences coalesce instead to the late Pleistocene, suggesting more recent nuclear gene flow between islands at low sea stands via forested connections.

Despite the overall similarity in phylogeographic patterns observed between mitochondrial and nuclear genomes, we observed several discordancies that provide the first indirect evidence for male-biased dispersal in colugos. In mammals, male-biased dispersal is common and often results in introgression of genetic markers with lower intraspecific gene flow, like the maternally inherited mitochondrial DNA, from one species into the nuclear genome of an invading species (Petit & Excoffier 2009). This pattern is highlighted by the strong difference in the position of Javan colugos between nuclear and mtDNA phylogenies. Nuclear timetrees indicate Javan colugos diverged ~4.0 mya from mainland, Peninsular Malaysian/Sumatran, and west Bornean colugos. However, in the mtDNA phylogeny Javan colugos are sister to east Bornean colugos, having diverged much earlier at ~9.3 mya (Fig. 3.3), suggesting an original colonization of Java from Borneo. During Pliocene glacial maxima, migrating colugo males likely

would have dispersed into Java from western Sundaic source populations, while capturing the local mtDNA genome that is more similar to East Bornean colugos. Similar scenarios of male-mediated nuclear gene flow would explain the much older mtDNA versus nuclear divergence times of Philippine colugos (Fig. 3.3).

To underscore the magnitude of population genetic differentiation within both colugo genera we calculated between-group ML genetic distances for populations represented in the biparental and mitogenomic phylogenies (Fig. 3.3) and compared them to genetic distances between well-established species (Fig. 3.4). The average between-group mtDNA genetic distance was 11.7% (sd=3.57%, min=5.8%), and the neutral X-chromosome distance was 0.56% (sd=0.13%, min=0.27%) between groups (tables B3.8-B3.9), exceeding divergences between numerous well-accepted primate and Sundaic species (Fig. 3.4)(14). Partitioning of mtDNA genetic variation was similarly high among seven Sundaic populations (FST=0.89, p<1e-05) (see methods) and little within populations (table B3.11). Philippine colugo mitochondrial DNA population differentiation between and within islands was also very high (FST=0.96, p<1e-05; 4.42% mean divergence), while the average X-chromosome divergence is 0.12% (sd=0.02%, min=0.1% between three sampled populations, tables B3.8, B3.9), with five mitochondrial and three nuclear lineages displaying equivalent or greater genetic divergence than is observed between many primate species (Fig. 3.4) (Brown et al. 2014). BPP analyses show the most significant support (PP>0.95) for six and two species (tables B3.16-B3.17). Considering these results we argue that Sundaic and Philippine lineages each comprise multiple distinct species based on the application of

modern species concepts (e.g., genetic, general lineage) that recognize separately evolving lineages (Baker & Bradley 2006).

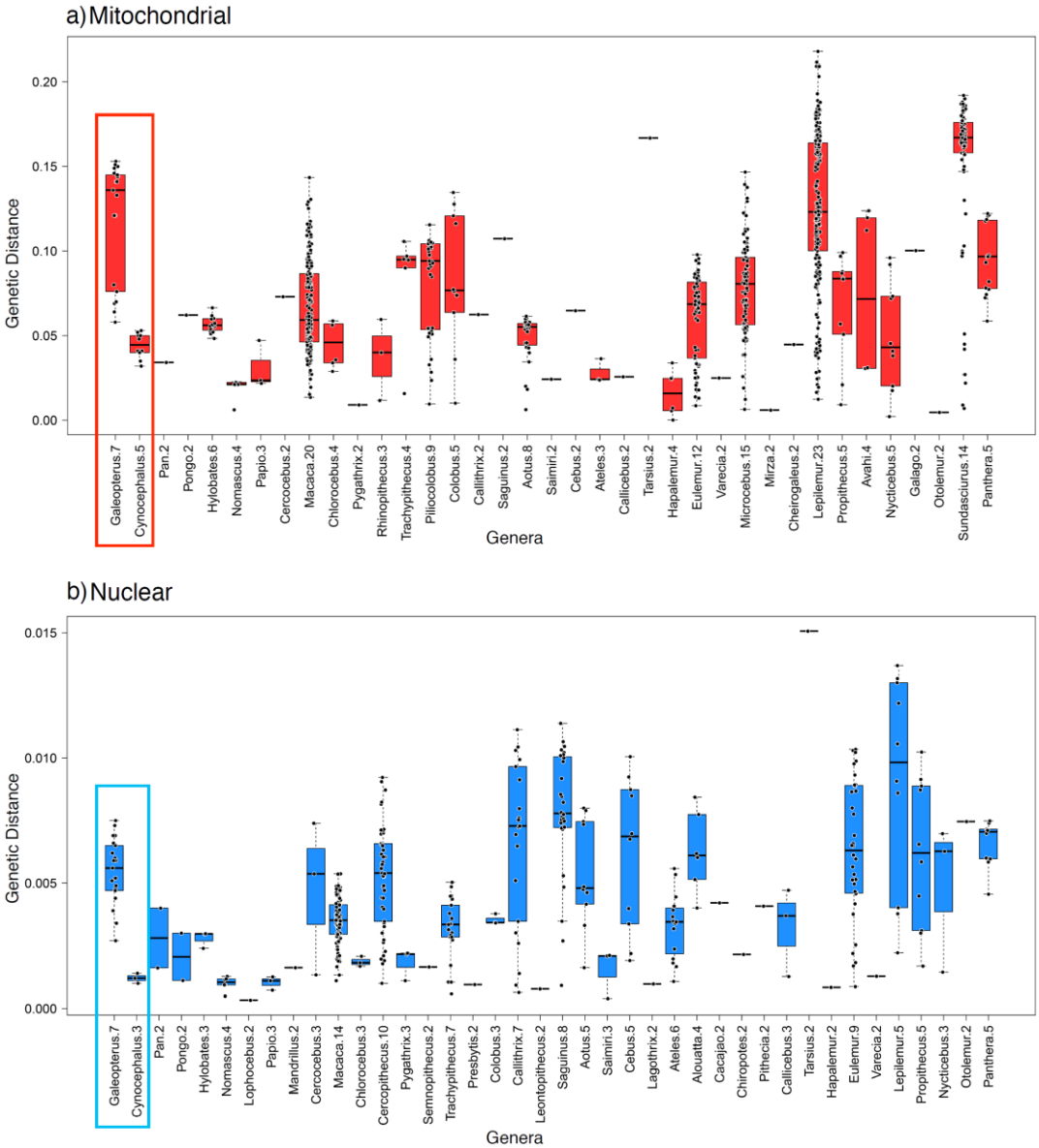


Figure 3.4. Comparison of genetic distance between well-established species (see methods) and proposed species groups for *Galeopterus* and *Cynocephalus.* A) mtDNA and B) nuclear DNA. The *x*-axis lists the name of each genus followed by the number of species in that genus that were compared. For colugo genera this number represents a conservative number of proposed groups/species. Colugo genera are highlighted in red and blue boxes.

Our findings have far-reaching conservation implications for the conservation of Sunda and Philippine colugos, which are presently listed as 'least concern' on the IUCN Red List when considered as just two species. Here we present concordant mitochondrial and nuclear genetic evidence for seven to eight colugo populations that should be recognized as evolutionary significant units (ESUs, Moritz 1994), or even distinct species, deserving of a conservation management strategy. Inclusion of additional, deeply divergent (i.e., >4-5%) mitochondrial lineages from populations that currently lack nuclear DNA data may increase the number of ESUs to fourteen (fig. A3.8, table B3.12). The current and future status of many of these smaller, isolated species level taxa (e.g., West Java) is uncertain given their present Red List status. By 2010 ~70% of the primary lowland forests within Sundaland had been cut down (Wilcove et al. 2013). Much of this land has been converted into oil palm and rubber plantations, and deforestation continues apace. Logging in the Philippines has also led to >90% reduction in forest cover over the past century (Brown & Diesmos 2009). Population and ecological assessments within Singapore report that while colugos can persist quite well within secondary tropical forests with >95% canopy cover, they are rarely found within plantation boundaries (Lim 2007, Lim et al. 2013). Therefore preserving minimally disturbed forests with high-density canopies within the range of these newly defined species will be critical for their future persistence, and may facilitate the survival of many other endangered species in this region.

3.3 Methods

3.3.1 G_variegatus-3.0.2 Genome Sample and DNA Extraction

The DNA used for sequencing the Malayan flying lemur, *Galeopterus variegatus*, was derived from a single male animal collected in West Java by Minoru Baba (Kitakyushu Museum of Natural History and Human History, Japan) under Indonesian Institute of Sciences Research Permits 6541/I/KS/1999, 3452/SU/KS/2002, and 3380/SU/KS/2003 (Janečka et al. 2008). Ethanol preserved tissue was used to extract genomic DNA with a Qiagen DNeasy Blood and Tissue Kit.

3.3.2 Genome Sequencing and Assembly

Total input sequence coverage of Illumina reads was 60× (45× 230-330 bp short inserts, 15× 3 kb mate pairs, and 5× 8 mate pairs) using a genome size estimate of 3.0Gb. The assembled sequence coverage was 55×. The combined sequence reads were assembled using the SOAPdenovo2 software (Luo et al. 2012). The assembly was improved using an unpublished program designed to close gaps, and SSPACE (Boetzer 2010). This draft assembly was referred to as Galeopterus_variegatus-3.0.2. This version has been gap filled, error-corrected with approx. 12X Illumina reads, and cleaned of contaminating contigs. The assembly is made up of a total of 179,513 scaffolds with an N50 scaffold length of 249 kb (N50 contig length was 20.8 kb). The assembly spans over 3.2 Gb.

3.3.3 RNAseq

Annotation for the G_variegatus-3.0.2 genome assembly was performed by NCBI and utilized RNAseq data from sequence read archive (SRA) samples

SAMN02736899 (*Galeopterus variegatus*), and SAMN02736900 (*Galeopterus variegatus*).

3.3.4 Constructing CDS Sequences and Determining Orthology

One-to-one orthologous amino acid (AA) and nucleotide (NT) raw alignments for nine taxa (*Homo sapiens, Pan troglodytes, Macaca mulatta, Callithrix jacchus, Otolemur garnettii, Ochotona princeps, Oryctolagus cuniculus, Canis familiaris, Felis catus*) (Ensembl v.79) were downloaded from OrthoMaMv9 (Douzery et al. 2014). Coding DNA sequences (CDS) were extracted from genome scaffolds for *Galeopterus variegatus* (GVA) and *Tupaia chinensis* (TCH) using their respecitve .gff annotation files. CDS sequences were constructed for *Cynocephalus volans* (CVO) using the GVA.gff annotation file and consensus sequences derived from aligning CVO reads to the GVA reference genome with BWA (settings= -n 0.001 -o 1 -l 24 -k 2). All CDS sequences were translated to AA sequences and only the longest isoform was kept for GVA, TCH, and CVO. Orthologous sequences were determined by a three-way protein BLAST of human, mouse, and dog AA sequences to the longest AA isoforms from GVA, CVO, and TCH followed by BLAST filtration. Filtration required each query taxon (human, mouse, dog) to have only one best BLAST hit (BBH) per database species (GVA or TCH or CVO), the query sequence to have had >50% of database sequence bases covered, and the BBH bit-score must be 2% greater than the second BBH bit-score.

3.3.5 Aligning Orthologous AA Isoforms to Existing OrthoMaMv9 AA Alignment

Unaligned orthologous AA longest isoforms from GVA, TCH, CVO were aligned to the existing raw AA alignment using Mafft (v7.127) with the --seed option (Katoh & Standley 2013). Any stop codon characters ('*' or '_') were masked with 'X' before alignment. This ensured proper reverse-translation of AA alignments to AA guided NT alignments. The stop codons were re-introduced in the AA guided NT alignments. Stop codon positions were recorded and any NT gene alignment with an internal stop codon not near the terminus was removed. Nucleotides corresponding to terminal stop codons were removed before analysis.

3.3.6 Genomic Alignment Filtration

AA genome wide CDS alignments were filtered by Gblocks v0.9.1 (Talavera & Castresana 2007), and subsequently filtered with two custom python scripts. The first removes the entire gene alignment if one or more individuals include frameshift mutations which result in poor alignments for the length of the gene. The second removes highly divergent windows, which is useful for masking poorly aligned isoforms, and highly divergent regions of dubious orthology. Filtrations were applied to AA alignments with window-size = 10AA, employing a pairwise divergence calculation between each taxon and human. Pairwise deletion was applied to divergence calculations within windows such that gaps and missing data were skipped and matched nucleotide pairs = the numerator and matched + mismatched nucleotide pairs = the denominator. Windows were excluded (masked with 'X's) if the AA divergence for one taxon was greater than two standard deviations away from the mean AA divergence across

windows from identical alignment coordinates assuming a normal distribution, and AA window must be greater than 20% diverged from human. In addition, all post-filtration AA sequences required >50% AA coverage and ≥20 AA in total length. Filtered AA sequences were backtranslated to nucleotide sequences using custom python scripts. Nucleotide sequences only contain bases corresponding to unambiguous (i.e., not 'X') amino acid bases, and skip blocks removed by Gblocks (Talavera & Castresana 2007).

We also applied a filter to identify and remove poor alignments and putative paralogous sequences. First, we constructed maximum likelihood phylogenies for each gene alignment. Genes were excluded if the phylogenetic distance between any two pairs of taxa was more than 2.5x the observed genetic distance between human and mouse sequences, or over 3.75 standard deviations away from the mean branch length between terminal nodes for the 21 taxon dataset. The same filter was applied to the 12 taxon dataset to estimate the divergence time between colugo genera, however we used a branch length cutoff of 2.5x the distance between human and dog as the criterion for gene exclusion. We chose human, mouse and dog for these calculations as the genome assemblies for these species are among the highest quality mammalian genomes, with relatively large phylogenetic distances that we could use to compare to other taxa in the tree.

3.3.7 Phylogenetic Analyses

RAxML v8.1.17 was used for all phylogenetic analyses with rapid bootstrap algorithm '-f a', GTR+gamma '-m GTRGAMMA', and 1000 bootstrap replicates for all nucleotide phylogenies (Stamatakis 2014). Amino acid phylogenies were constructed

with similar settings and JTT amino acid substitution matrix + gamma '-m PROTGAMMAJTT'.

3.3.8 Coding Gene INDELs

A pool of potentially phylogenetically informative indels were identified by custom python script that searches AA gene alignments with indels supporting a specific phylogenetic hypothesis. Input gene alignments were unfiltered whole genome AA coding sequence alignments (OrthoMaMv9) with *Galeopterus variegatus*, *Cynocephalus volans*, and *Tupaia chinensis* synthesized AA coding sequences added to alignments as described in supplemental methods 2.2. To identify a deletion this code looks for shared gaps in taxa specified for in particular hypothesis, and then the opposite for insertions. Deletion examples for Primatomorpha (*Homo, Pan, Gorilla, Pongo, Nomascus, Papio, Macaca, Callithrix, Tarsius, Otolemur, Microcebus,* and *Galeopterus*) were specified as taxa that should have a shared gap of equal length in a sequence alignment. We relaxed indel identification to only require a subset of taxa (*Homo, Macaca, Galeopterus*) to have the deletion while the rest of the taxa (*Pan, Gorilla, Pongo, Nomascus, Papio, Callithrix, Tarsius, Otolemur, Microcebus*) could have the deletion, but were not required to have the deletion. We relaxed indel identification requirements to only require a subset of taxa to match an indel because some taxa had alignment errors which would prevent identification of the indel if all taxa were required to harbor the indel. *Homo, Macaca,* and *Galeopterus* were required to support primatomorpha indels, *Galeopterus, Tupaia chinensis* and/or *Tupaia belangeri* were required to support Sundatheria indels, and *Homo, Macaca,* and *Tupaia chinensis* were required to support

Primate+Scandentia indels. Indels were removed from consideration if any unspecified taxa had a deletion (homoplasy, or alignment error) with the same coordinates as shared deletions from the taxa specified in the hypothesis. After indels in gene alignments were identified that contain potentially informative indels, gene alignments were manually curated to identify phylogenetically informative indels (figure A3.4). We identified indels flanked by conserved AA sequence and that were present in all taxa for a particular hypothesis. Chi-square calculations are shown in table B3.2, and followed (Waddell 2001).

3.3.9 Retrotransposons

Given the newly sequenced colugo genome, we further explored the question of how closely flying lemurs are related to primates by focusing on more complex, very reliable ancient changes and in considering all possible phylogenetic scenarios by screening for and analyzing the integration patterns of retroposed elements, virtually non-homoplastic phylogenetic markers. In mammals, retrotransposons integrated continuously over time, and were accompanied by duplications of randomly selected, 4-30 nt (4-12 for LTRs and 8-30 for LINE1s) of the coincidental genomic target sites. Target site duplications enable verification of orthology of diagnostic retroposon insertions in different taxa. Identical retroposon insertions in two species and an orthologous empty site in a third species supports the monophyly of the two and provides no support for relatedness of the third. We systematically tested all possible evolutionary hypotheses relating colugos to treeshrews (Scandentia), and human (primates), by statistically considering the following three possible evolutionary

scenarios: a phylogenetic group composed of 1) colugo plus human, 2) colugo plus treeshrew, or 3) treeshrew plus human.

Based on a previous successful screening in Euarchontoglires, we searched the newly sequenced colugo genome (figure A3.3) for Long Terminal Repeats (LTR, MLT1A/MSTD) and Long INterspersed Element (LINE, L1MA5/6) retrotransposon subfamilies, which were both active during the euarchontan speciation (Kriegs et al. 2006, Kriegs et al. 2007). We then compared the 29,222 LTR-MLT1A and 12,983 LINE-L1MA5/6 hits we received along with their flanking target site duplications to other euarchontan genomes, which yielded 221 pairwise-aligned regions. After comparing these to the genomes of additional outgroup species (pika, rabbit, mouse, kangaroo rat, guinea pig, squirrel, dog, cat, megabat, horse, elephant, rock hyrax, sloth, and armadillo), we generated a retroposon presence/absence pattern for these species. Seventeen of these retroposons were phylogenetically informative; 12 LTR-MLT1A/MSTD elements and 4 LINE-L1MA5/6 elements were present in both colugo and human but were absent in tree shrew and other mammals. One additional LTR-MSTD element was present in both colugo and tree shrew but absent in human and outgroups. To specifically test the 3rd hypothesis (treeshrews+primates), we also screened 66,860 loci of the 2-way (UCSC) alignment human-*Tupaia* (34,703 MLT1A/MSTD and 32157 LiMA5/6), yielding 198 orthologous elements in human and tree shrew, but none were found that were also absent in colugo and outgroup species (figures A3.2-A3.3). Sixteen retroposon elements shared between human and flying lemur support the sister-group relatedness of these two eutherian orders. The KKSC

statistical test for genomic insertion data (http://retrogenomics.uni-muenster.de:3838/KKSC_significance_test/) was significant (p=2.7e-07).

3.3.10 Sensory Gene Family Expansions and Positive Selection

Published *V1R* and *OR* gene sequences from human, mouse, rat, cow, dog, and opossum were used as the query sequences for BLAST searches against the domestic cat genome. We enforced an *E*-value threshold of 10^{-5} for filtering BLAST results. All identified sequences were extended 1.5Kb on either side for open reading frame identification and assessment of functionality. If multiple start codons were found, the alignment results of known intact mammalian V1R and OR amino acid sequences were used as guidance. Any putative genes containing early stop codons, frameshift mutations, and/or incomplete gene structure (i.e., 3 extracellular regions, 7 transmembrane regions and 3 intracellular regions) were designated as pseudogenes. To confirm orthology, we aligned all members of the *V1R* and *OR* gene families and constructed maximum likelihood trees. We compared the *V1R* and *OR* gene trees generated above to a mammalian species tree (Meredith et al. 2011) to estimate gene gain and loss using the software Notung (Chen et al. 2000).

Two datasets were constructed to test for positive selection: **1)** a seven-taxon dataset (*Homo sapiens, Callithrix jacchus, Otolemur garnettii, Galeopterus variegatus, Tupaia belangeri chinensis, Mus musculus, Canis familiaris*) and an eight-taxon dataset (*Homo sapiens, Callithrix jacchus, Otolemur garnettii, Galeopterus variegatus, Cynocephalus volans, Tupaia belangeri chinensis, Mus musculus, Canis familiaris*). Amino acid sequences were downloaded from OrthoMaMv9. The initial seven-taxon

dataset contained 8,514 gene orthologs, and the eight-taxon contained 4,899 genes. Individual genes were removed if at least one taxon possessed a frameshift mutation or pre-mature stop codon. Sequences were back translated to nucleotide sequences. Genes that contained one-to-one orthologs for all seven or eight taxa were aligned with Prank (Löytynoja 2014) and filtered as described in section 2.3. A PERL script pipeline was applied which removed poorly aligned or incorrectly annotated amino acid residues caused by obvious gene annotation errors within the genome assemblies. Aligned amino acid sequences were used for guiding nucleotide coding sequences by adding insertion gaps and removing poorly aligned regions. We estimated nonsynonymous and synonymous substitution rates using the software PAML4.0 (Yang 2007). We used both branch-site and branch models as described (Montague et al. 2014) to identify accelerated rates of genes on specific branches of an evolutionary tree, and specific amino acid residues that were potentially under positive selection. Paired models representing different hypotheses consisted of branch tests, and branch-site tests (fixed ω=1 vs. variable ω). For the branch-specific tests, free ratio vs. one ratio tests were used to identify putatively positively selected genes. These genes were subsequently tested by two ratio and one ratio models to identify genes with significant positive selection of one branch versus all other branches (two branch test). Significance of LRT results employed a threshold of $p<0.05$. We assessed enrichment of KEGG pathway and disease gene association tests using Webgestalt (http://bioinfo.vanderbilt.edu/webgestalt/). Gene symbols were used as input, and the organism of interest setting=*Homo sapiens*. Only significant KEGG Pathways and Disease Association categories were reported, using a

hypergeometric test and the significance level set at 0.05, implementing the Benjamini and Hochberg multiple test adjustment to control for false discovery. The most significant enrichment for genes in the colugo lineage were related to various categories of cardiovascular disease and lipid metabolism in humans, notably those encoding apolipoproteins (i.e. *APOE* and *APOH*) that function in phospholipid and lipoprotein metabolism (Fig. 3.2c). However, we argue that this enrichment category was likely driven by a large number of genes with pleiotropic effects in both sensory systems and skeletal-muscular function (Fig. 3.2c; tables C3.1-C3.4).

3.3.11 Colugo Population Sampling and DNA Extraction

Ethanol-preserved tissues for *Cynocephalus volans* were obtained from the Field Museum of Natural History (Chicago, IL, USA). DNA was extracted using Qiagen DNeasy Blood and Tissue Kit following manufacturers specifications.

Dried museum tissues were sampled from three different institutions: the National Museum of Natural History (NMNH, Smithsonian), the American Museum of Natural History (AMNH), and the Raffles Museum of Biodiversity Research (RMBR). For the majority of specimens we removed ~5mg of adherent tissue from inside the cranial cavity or nasal turbinate system. For some specimens we collected multiple sample types, including hair, skin, cartilage, and bone. DNA was extracted from all tissues through proteinase K digestion, protein precipitation and removal, and ethanol precipitation of DNA. Digestion was performed with 520µl Cell Lysis Solution (Puregene® D-5002, Gentra, Qiagen), 600µg Proteinase K, and 50µg linear acrylamide (Ambion) and incubated in rotating heat block at 60°C for 48 hours. Undigested samples

were disrupted with a pestle after 24 hours. Protein precipitation and ethanol DNA precipitation followed the standard guidelines of the Gentra/Puregene DNA isolation protocol (Qiagen). Sample information is described in table C3.5.

3.3.12 Illumina Library Preparation, Low Coverage Sequencing, and Nuclear Capture

Illumina libraries for low-coverage sequencing were prepared with the Illumina Truseq HT Dual Indexing kit following manufacturer's specifications, except that Microcon-30 centrifugal filters (Millipore) were used following the blunt-ending step, prior to adapter ligation, to retain degraded, low MW DNA fragments (~≥50-bp). Illumina libraries for nuclear capture were prepared with the Nextflex Rapid DNA Sequencing Kit following manufacturer's specifications. Libraries were sequenced on the Illumina HiSeq 2000.

Nuclear capture was performed following Mason et al. (Mason 2011) with slight modifications, including a 72 hour hybridization reaction with Illumina adapter blocking oligos (Del Mastro & Lovett 1997, Maricic et al. 2010). Primers for probe amplification were designed from the G_variegatus-3.0.2 genome assembly (table C3.5). Capture probes were generated through PCR amplification of ~1kb DNA fragments from modern *Galeopterus* and *Cynocephalus* DNA extracts (Janečka 2008). Four separate probe pools were generated, including individuals from different geographical locations. These were applied to different target samples, based on locality, to minimize sequence divergence between the probe and the sample during capture experiments. Probes were amplified from high molecular weight DNA extracted from the following frozen tissue samples:

GVA_03 (Singapore, Peninsular Malaysia), GVA_04 (West Java), CVO_02 (Leyte, Philippines). Three probe pools (1. Peninsular Malaysia, 2. West Java, 3. Peninsular Malaysia + West Java) were used to perform hybrid capture from *Galeopterus* samples, and the CVO_02 probes were used to perform hybrid capture from *Cynocephalus* samples. Amplifications were performed with Platinum-Taq DNA polymerase (Invitrogen), 1.5mM $MgCl_2$, 0.8mM dNTPs, 2μM primers, under the following cycling conditions: 2 minutes hot start at 94°C, denaturation for 30 seconds at 94°C, touchdown annealing at 2 cycles each of 60°C, 58°C, 56°C, 54°C, 52°C, followed by 30 cycles at 50°C, extension for 1 minute at 72°C, and final extension for 5 minutes at 72°C. Successful amplicons were pooled (equal volume) and labeled with biotin using Biotin-High Prime (Roche).

3.3.13 Probe Design and Generation

X-chromosome probes: We queried the draft colugo genome assembly using a set of human and mouse 1:1 orthologous X chromosome coding sequences (Ensembl v67) using BLAST, to identify candidate colugo X-chromosome contigs. We then performed reverse-BLAST of the top scoring colugo contigs back to the human (GRCh37.p7), mouse (NCBI m37), and dog (CanFam 2.0) genome assemblies. Colugo contigs that had top scoring BLAST hits to the X-chromosome sequence from all three species were considered orthologous. We generated nearly neutral X chromosome capture probes by designing PCR-based amplicons ~1-kb in length from within each retrieved scaffold. We specifically targeted non-repetitive sequence that was the greatest possible distance from any annotated coding sequence within the scaffold.

Y-chromosome probes: Human and mouse single copy Y-chromosome genes (Skaletsky et al. 2003, Soh et al. 2014) were queried against the draft assembly of the *Galeopterus* genome using BLAST. We then performed reverse-BLAST of the top scoring colugo contigs back to the human genome (GRCh37.p7). We selected those contigs with either a best BLAST hit to the same single copy Y-chromosome sequences and/or to the X-chromosome with >15% sequence divergence. We also constructed maximum likelihood trees with the candidate contig and annotated X and Y orthologs to validate reciprocal monophyly of X and Y orthologous sequences. We generated nearly neutral Y-chromosome capture probes in the same manner as described above for the X chromosome probes. Primers were validated as Y-specific by simultaneous PCR screening on male and female DNA samples.

Autosomal gene probes: We designed capture probes to target a subset of selected protein coding genes that influence vision, coat color, and body size in mammals. Candidate human protein coding genes were queried against the draft assembly of the *Galeopterus* genome using BLAST. Identified exons were aligned to human genes and trimmed to human exon boundaries. Orthology vs. paralogy was determined by maximum likelihood phylogenetic tree construction, using a sequence matrix of known mammalian orthologues, as well as closely related paralogs. We selected all exon-containing contigs that formed a monophyletic group with orthologous mammalian exons for the genes of interest. Primers were designed for all targeted sequences with BatchPrimer3 (You et al. 2008).

3.3.14 Modern DNA Sequence Trimming and Filtration

Illumina sequences were filtered with TrimGalore! v0.3.3 to remove Illumina adapter sequences and trim low quality bases (Parameters: --paired --retain_unpaired -q 20 --length 30 --stringency 1 --length_1 31 length_2 31) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).

3.3.15 Modern Genome Reference Assemblies

We generated ~14X paired-end Illumina read coverage from a 300-bp avg. insert size Illumina library created from DNA of *Cynocephalus volans* **(Museum accession and SRA Accession number).** A reference assembly was constructed by aligning quality filtered Illumina sequences to the G_variegatus-3.0.2 genome assembly with BWA v0.7.5a-r405, bwa-mem, (http://bio-bwa.sourceforge.net/bwa.shtml). Approximately 5X coverage of pen-tailed treeshrew (*Ptilocercus lowii*) Illumina reads (**Reference and SRA Accession number)** were reference aligned to the *Tupaia chinensis* genome scaffolds (Fan et al. 2013) using bwa-mem.

3.3.16 Museum DNA Sequence Trimming and Filtration

Raw Illumina sequences were filtered with SeqPrep to remove Illumina adapter sequences, trim low quality bases, and merge overlapping sequence pairs (Parameters: -A AGATCGGAAGAGCACACGTC -B AGATCGGAAGAGCGTCGTGT -q 13 -o 15 -L 30 -g) (https://github.com/jstjohn/SeqPrep). We removed the first and last three bases of each sequence read, as these bases are highly susceptible to chemical damage (table B3.#) (Dabney et al. 2013).

3.3.17 mtDNA, Biparental, and Y-Chromosome Sequence Assembly

Mitochondrial genomes were assembled using both *de novo* (SOAP, CAP3 http://seq.cs.iastate.edu/cap3.html) (Luo et al. 2012, Huan & Madan 1999) and reference-based (BWA aln v0.7.5a-r405) (Li & Durbin 2009) assembly strategies. Multiple assemblies based on a range of k-mer values were performed on sequences from each individual. *De novo* assemblies that produced complete mitogenomes were used as reference sequences for other individuals, and selected based on geographic proximity, to reduce sequence divergence between the reference sequence and the assembled reads: three complete *Galeopterus* mtDNA genomes from GenBank (AJ428849.1, JN800721.1, AF460846.1) and five *de novo* mitogenome assemblies from samples GVA_22 (Palembang, Sumatra), GVA_45 (Sabah, Borneo), GVA_49 (Pulau Sebuko, Borneo), CVO_06 (Samar Island, Philippines), and CVO_08 (Tupi, Mindanao). Sequence reads from each sample was aligned to reference mitogenomes from several of the geographically closest candidates, with BWA aln parameters -n 0.0001 -o 1 -l 24 -k 3. The assembly with the highest percentage of reference bases covered and highest average depth of sequence was chosen for final consensus sequence generation.

Biparentally-inherited target loci were extracted from the G_variegatus-3.0.2 genome assembly, and used as the reference sequence for reference-based assemblies for all *Galeopterus* assemblies. Sample CVO_08 was chosen as the reference sequence for all *Cynocephalus* assemblies because it mapped to the G_variegatus-3.0.2 probe sequences with the highest reference base percent coverage and average depth.

Reference assemblies were mapped using BWA with parameters (-n 0.001 -o 1 -l 24 -k 2).

Y chromosome sequence alignments for male Sunda colugos were constructed by aligning to Y chromosome scaffolds identified within the G_variegatus-3.0.2 genome assembly. Philippine colugos were aligned to a *Cynocephalus* reference-based Y chromosome consensus sequence generated from aligning the reads of the highest quality male DNA specimen, CVO_10, to G_variegatus-3.0.2 Y chromosome scaffolds.

3.3.18 Consensus Sequences

Consensus sequences were called using SAMtools (Lit et al. 2009)(v1.1-26-g29b0367 (htslib 1.1-90-g9a88137)) mpileup, vcftools (v1.1-88-g4c0d79d (htslib 1.1-90-g9a88137)), and vcfutils.pl. All biparental sequences were called as diploid requiring a minimum read depth=3. Y chromosome sequences were called as haploid and required a minimum depth=2.

3.3.19 Sequence Alignments and Phylogenetic Analyses

Mafft v7.127 (Katoh & Standley 2013) was used to align all consensus sequences. Alignments were manually curated to remove poorly aligned regions. RAxML v8.1.17 was used for all phylogenetic analyses with rapid bootstrap algorithm '-f a', GTR+gamma '-m GTRGAMMA', and 1,000 bootstrap replicates for all nucleotide phylogenies and 'm PROTGAMMAJTT' for all amino acid phylogenies (Stamatakis 2014).

3.3.20 Genetic Distance Estimates

MEGA6 v6.06 (Tamura et al. 2013) was used to calculate between-group mean genetic distances. Groups were defined as strongly supported (100%) monophyly groups (when more than one individual) which possessed at least ~3-5% between-group mtDNA divergence, and in some cases, corroborative nuclear phylogenetic structure. MEGA-CC v7.0.7 (Kumar et al. 2012) was used to calculate between-species divergence levels within genera sampled from Perelman et al. 2011. Species comparisons were required to have >20% total sequence coverage for mitogenome data, and >50% coverage for nuclear data. We used the maximum composite likelihood distance with a gamma shape parameter = 4, and pairwise deletion (removing all ambiguous bases for each sequence pair) for all distance calculations. We required a minimum of two species per genus for comparisons to other primate and Sundaic mitochondrial and nuclear gene sequences. *Panthera* interspecific distance calculations were derived from whole genome alignments between all species (Li et al. 2016). Boxplots were constructed in R v3.1.2 (R Core Team 2014).

3.3.21 Analysis of Molecular Variance (AMOVA)

To estimate the degree of differentiation among populations, we estimated fixation indices (F_{ST}) using an analysis of variance (AMOVA) approach calculated in Arlequin v3.5.2.2 (Excoffier & Lischer 2010). Mitochondrial haplotype data was compiled in DNAsp v5 using a complete deletion-option alignment (7,176bp final sites) for 45 Sunda colugos to make the input Arlequin file (Librado & Rozas 2009). We defined seven populations for the AMOVA, with 16,000 permutations.

3.3.22 Principal Component Analysis (PCA) of Genetic Variation

We analyzed 1,340 SNPs present in our X-chromosome capture data from 12 individuals that represent the 7 major clades of colugos present in the Biparental phylogenies. Specimen read-group (@RG) and sample (SM:) information was introduced to each .bam alignment file during the alignment. SNPs were called through samtools mpileup, bcftools, and vcfutils.pl. SNPs were required to have a minimum root mean squared mapping quality of 30, minimum depth of 3, and maximum depth of 100.

We used the R-package SNPRelate to perform the PCA from the.vcf file. The first five principal components explain 19.4, 13.7, 12.2, 10.9, and 9.1 percent of the total variation. Therefore only 65.3% of the total X-chromosome SNP variation is explained by the first five principal components. Even though only a small subset of the variation can be explained in two-dimensional space we still see clear separation of 5 of the seven proposed species in the first two principal components (figure A3.15). Biplots for principal components 1-6 are shown in figures A3.15-A3.19. The biplot for PC1 vs PC2 illustrates almost no variation between Peninsular Malaysian/Sumatran individuals and the W. Bornean individual (GVA_16). However, the first two principal components are inappropriate for comparisons to GVA_16, because GVA_16 is only minimally correlated with PC1 (r^2= 0.11) and PC2 (r^2= -0.05) and therefore does not vary along dimensions described by PC1 and PC2. On the other hand GVA_16 is highly correlated with PC5 (r^2= -0.35) and PC6 (r^2= -0.73) (table B3.16) meaning GVA_16 varies along these axes and the total variance explained by PC5 and PC6 is still substantial: 9.1% and 7.5% respectively, indicating that a larger proportion of variance could be explained by

PC5 and PC6 than by PC1 and PC2 for GVA_16. Because the total variance explained by each principal is known and the sum of squares of one principal component's loadings is equal to 1, we can calculate the proportion of variance explained by each variable (individual in this case) for each principal component. $v_i = (l)^2*(v)$ where v_i = percent of total variance explained by one variable (individual) for one principal component, l = variable component loading, and v = percentage of total variation explained by this one principal component. Therefore we calculate that the percentage of total variation explained by GVA_16 for PC1 and PC2 to be equal to 0.23% and 0.03%, while for PC5 and PC6 the percentage of total variation is 1.15% and 4.06%, respectively.

3.3.23 Divergence Dating Calibrations

The poor dermopteran fossil record precludes the application of internal fossil calibrations. Therefore we calibrated these phylogenies with the 95% confidence intervals of the estimated *Galeopterus-Cynocephalus* divergence date, estimated from the genome-wide orthologous CDS matrix and bounded with seven non-dermopteran fossil calibrations (Meredith et al. 2011) (table B3.7). We estimated the divergence time between *Galeopterus* and *Cynocephalus* from 2,729 genome-wide orthologous coding gene alignments (final matrix length=3,515,409 bp) extracted from 12 mammalian genomes (*Homo sapiens, Pan troglodytes, Macaca mulatta, Callithrix jacchus, Otolemur garnettii, Oryctolagus cuniculus, Ochotona princeps, Canis familiaris, Felis catus, Galeopterus variegatus, Cynocephalus volans, Tupaia chinensis*) (OrthoMaMv9)(Douzery et al. 2014). We employed seven external fossil calibrations

(table B3.7) as minimum and maximum constraints (Meredith et al. 2011). Divergence time estimation was performed under several different analysis conditions that varied both rates (independent and autocorrelated) and calibration (hard and soft) following Meredith et al. (Meredith et al. 2011). We chose an approximate-likelihood method under a GTR+Γ model of sequence evolution in the MCMCTree package (Yang 2007).

3.3.24 MCMCTree

All divergence dating was completed with MCMCTree v4.8a within PAML (Yang 2007). We used v4.8a, which implements a revised dirichlet prior, enabling proper estimation of rgene_gamma (substitution rate through unit time) and proper retention of uncertainty in confidence intervals from fossil calibrations (Dos Reis et al. 2014). All MCMCTree calculations were performed twice to ensure convergence. Time-tree branch lengths were averaged from multiple runs, and the maximal range of 95% confidence intervals were kept to represent maximal uncertainty for each node.
The rgene_gamma prior shape and scale (α and β) values were estimated by first calculating the clock-like substitution rate per unit time in baseml for the whole phylogeny given a nucleotide alignment and rooted phylogeny with branch lengths, and point estimates of divergence time at available calibrated nodes. Prior values α = (m/s)^2 and β = m/s^2 where m = mean and s = standard deviation. Mean = standard deviation of the gamma distribution when the shape parameter α = 1. When α = 1 then we solve for β with m = s.

3.3.25 Molecular Divergence-time Estimate Between Colugo Genera

The mean divergence time between *Galeopterus* and *Cynocephalus* derived from these analyses was estimated as 11.3-Mya (with a 95% credibility interval of 5.2-19.6 Mya). The point estimate coincides approximately with the lowest sea stand of any prior to that during the Tertiary (Meijaard 2004), and represents the first glacial period that lowered sea levels below present day levels during the entire Miocene (Haq 1987, Uba et. al. 2007). We note, however, that conflicting long-term and short-term eustatic sea level curves have been reported throughout the Neogene (see Kominz 2008). We hypothesize that a forested corridor between Borneo and the Philippines must have been present to facilitate colonization of the Philippines, likely via a route formed along the current Sulu archipelago**.**

3.3.26 Molecular Divergence-time Estimates Within Colugo Genera

We estimated divergence times within each colugo genus using MCMCtree, assuming an autocorrelated rates model with a soft calibration for the basal split between *Cynocephalus* and *Galeopterus* derived from the 95% confidence intervals of our molecular supermatrix-based estimate (Supplement section 4.3). Calculations were performed with exact likelihood and an HKY-85+Γ model of sequence evolution. Datasets were reduced to one taxon per divergent lineage, selecting individuals with the greatest capture probe coverage.

3.3.27 Craniodental Morphometric Analyses

Morphometric data included 19 linear craniodental measurements taken from 82 Sunda colugo skulls after sampling tissue from museum specimens (table C3.7). Data

was log normalized before principal component analyses (PCA). PCA analyses were performed with the R package ‘prcomp’ by singular value decomposition. We conducted PCA with and without normalizing for body size (fig. A3.13). Condylobasal length (CBL) is a measurement of skull length and is correlated with body size, we therefore normalized measurements by body size, dividing by CBL (Stafford & Szalay 2000).

We observed the most geographic sorting in PCAs without normalizing for body size and when recently diverged dwarf individuals were removed (fig. A3.13). After males, females, and dwarfs are normalized by body size we see little geographic structuring, indicating that most of the variation in craniodental measurements is due to body size variation and confirms observations of Stafford and Szalay (2000) (fig. A3.13-A3.14). This is expected as all measurements are highly correlated with CBL (mean=0.68, standard deviation=0.18). However, the measurement of ‘min.w.temps’ is least correlated with body size (r=0.26). The ‘min.w.temps’ (the minimum distance between the temporal lines on the roof of the skull) vector is of significant magnitude and tends to sort Bornean colugos to a subset of the distribution after body size normalization (fig. A3.14, table B3.15). No PCAs based on morphology were capable of sorting colugo populations to mutually exclusive clusters, however they did generally sort based on regional geographic distribution.

Dwarf colugos were defined as individuals with a 10% reduction in CBL compared to the average CBL of neighboring populations from large islands or the mainland. Dwarf individuals residing on satellite islands represent recent deviations in phenotype when compared to the morphology of larger islands, and therefore are not

representative of the deeper evolutionary history of Sunda colugo species. This again agrees with Stafford and Szalay (2000), who concluded that dwarf populations did not warrant species-level classifications based on body size reduction alone.

3.3.28 Species Classifications, BPP and Conservation Units

The genetic species concept (Baker & Bradley 2006) argues that species can be classified based on genetic isolation rather than reproductive isolation. We present evidence for multiple, genetically divergent populations within the Sunda and Philippine colugo that in the majority of cases are consistent across two or more genetic transmission components (*i.e.* mtDNA, Y chromosome, and biparentally inherited loci). The genetic divergence levels between seven colugo populations exceed those between numerous well-established species within other mammalian orders. Nuclear and mtDNA genetic divergence levels conservatively support a classification scheme that recognizes a minimum of seven species within *Galeopterus*. Three additional evolutionary significant units (ESUs) (Moritz 1994), and potentially valid species, may be recognized within the genus: SW. Borneo (GVA_58, GVA_61), SE Borneo (GVA_49) and S. Sumatra (GVA_21, GVA_22, GVA_28) based on divergent (>4%) mitochondrial haplotypes.

The nuclear sequence divergence estimated for all *Cynocephalus* pairwise comparisons exceeded that of at least seven pairs of described Primate species (Fig. 3.4), validating a minimum of three species level taxa within the Philippines: Leyte, Dinagat, and eastern Mindanao. Only mtDNA was obtained from specimens of colugos sampled from western Mindanao (Zamboanga Peninsula), Basilan, and Bohol. Mitochondrial

divergence between Basilan and western Mindanao was less than 1%. However, the mtDNA sequence divergence between Bohol and western Mindanao and between these two populations and all other Philippine populations was between 3.2 and 4.1% (est. divergence time $\geq$1.5 Mya) (table B3.8, Fig. 3.3A). Given the general concordance between divergent nuclear and mitochondrial lineages, we consider each of the following five Philippine populations as evolutionary significant units worthy of formal subspecific (and possible species-rank) recognition and separate conservation strategies: 1) eastern Mindanao, 2) western Mindanao + Basilan, 3) Dinagat, 4) Leyte, and 5) Bohol.

We used BP&P (Bayesian Phylogenetics and Phylogeography) (Yang 2015) to test our proposed colugo species groups against alternative species models. We provided a fixed guide tree based on the structure recovered from phylogenies constructed in RAxML (A10: speciesdelimitation = 1 and speciestree =0). The rjMCMC algorithm was used for species delimitation where both species delimitation = 1 0 2 and species delimitation = 1 1 2 1 were defined in the control file. . Equal prior probabilities for rooted trees was specified by speciesmodelprior = 1. The provided fixed phylogeny represented the seven proposed species groups of the Sunda colugo '((E Borneo, NE Borneo),((Vietnam, Laos),(Java, (West Borneo, Peninsular Malaysia+Sumatra))));' or the three proposed species of Philippine colugo '(Mindanao, (Leyte, Dinagat));', which were based on the three populations for which we successfully captured adequate nuclear DNA.

We estimated the gamma priors theta (θs) and tau (τs). A gamma prior's distribution is defined by two parameters shape parameter (α), and rate parameter (β). The shape parameter changes with how accurately the prior represents the data. High confidence prior values might have a high α value while low confidence priors should have a low α value. Increasing α restricts the gamma distribution reducing how much parameters can vary in the posterior, while low α results in a diffuse gamma distribution where estimated values can vary more freely in the Bayesian posterior (Yang 2015). To lessen restrictions on parameter estimates in the posterior we chose a diffuse shape parameter (α) = 2. We estimated the rate parameter β for gamma priors theta (θs) and tau (τs) with α = 2, the mean (m) of the gamma distribution, and the standard deviation (s) of the gamma distribution. Gamma prior theta is based upon the population size and the mean of the gamma distribution is calculated as the average proportion of different sites. The average between group genetic divergence for Sunda colugos is ~0.5% therefore the average proportion of differing sites is ~0.005 which is equal to the mean (m) of the gamma distribution. For *Cynocephalus* m = ~0.001. The relationship between the mean and standard deviation of a gamma distribution changes as α changes. The standard deviation of the gamma distribution is $s = m/\sqrt{\alpha}$. For *Galeopterus* $s = 0.005/\sqrt{2} \sim= 0.0035$. The mean and standard deviation are used to calculate β = m/s^2 = 0.005/0.0035^2 = 408. Calculated α and β for thetaprior = 2 408. The mean for tau was calculated as the years of divergence of the root of the tree divided by the mutation rate ($1x10^{-9}$). The mean of the gamma distribution for tau for *Galeopterus* = $5.5x10^{6}/1x10^{-9}$ = 0.0055. Divergence time for the root was derived from the biparental timetree which is

~5.5Mya. The same procedure was followed for calculating priors for tau and *Cynocephalus*. Estimating priors followed (Yang 2015) and the BPP documentation. All BPP runs were executed twice to confirm convergence. The sensitivity of BPP analyses was also assessed by varying the rate parameter (β) for theta and tau priors, following (Yang 2015). We used the biparental matrix for *Galeopterus* and *Cynocephalus* after complete deletion of all columns containing missing data. The *Galeopterus* dataset was sensitive to variations of β for theta when varied from 10 to 1000, however was not sensitive to variation in β for tau (table B3.16). The sensitivity of the biparental dataset for *Galeopterus* only changed the support values between the six and seven species models and the probability for the presence of a node separating Laos from Vietnam, however only six species were strongly supported, with posterior probability >0.95 (table B3.17). The *Cynocephalus* dataset was neither sensitive to variation in β for theta, when varied from 1000 to 2040, nor to variation in β from 10 to 1000 for tau (table B3.17).

3.3.29 Biogeography Notes

The presence of a north-south savannah corridor running through the South China Sea and the Javan Sea during Pleistocene glacial maxima would have likely prevented dispersal of forest-dependent taxa like colugos between Borneo and Sumatra+Peninsular Malaysia, and between Borneo and Java. However, the extent to which this savannah corridor was present, and the continuity of the corridor itself across its proposed distribution, is debated (Bird et al. 2005, Cannon et al. 2009, Raes 2014, de Bruyn 2014, Sheldon 2015). Glacial maxima generally are characterized as dry periods

with less precipitation, accompanied by drastically lowered sea levels (-120m) which expose sandy sea-bed soils (Slik 2011). Simulations have predicted that the last glacial maximum (LGM) was very dry and cold, suggesting the possibility of a continuous savannah corridor. This is supported by genetic evidence from many forest-dependent vertebrate species distributed between Borneo, Peninsular Malaysia and Sumatra, which possess estimated divergence times predating the LGM (Leonard 2015). However, even if there was a savannah corridor present at the LGM, there were many interglacial periods during prior millenia when forested corridors likely would have existed to connect these present-day landmasses (de Bruyn 2014, Cannon 2009, Raes 2014).

CHAPTER IV

COLUGO BIOGEOGRAPHY REVEALS PALEO-FOREST CORRIDORS THROUGHOUT SUNDALAND

4.1 Introduction

Determining past geographic distributions of forests is important for biogeographic inferences regarding the historical development of the current occurances and observed genetic variation in forest-dependent taxa. Many geological (Hall 2013), climatic (Cannon et al. 2009), environmental (Cannon et al. 2009, Raes et al. 2014), and biological (de Bruyn et al. 2014) indicators have been utilized to predict paleo-forest distributions. Here we focus on six mammalian species that serve as biological indicators, spread throughout the Southeast Asian mainland and archipelago: 1) the Sunda colugo *Galeopterus variegatus* (*G. variegatus*), 2) the Philippine colugo *Cynocephalus volans* (*C. volans*) 3) the lesser mouse deer *Tragulus kanchil* (*T. kanchil*), 4) the greater mouse deer *Tragulus napu* (*T. napu*), 5) the Javan mouse deer *Tragulus javanicus* (*T. javanicus*)(Meijaard & Groves 2004) 6) the Sunda pangolin *Manis javanica* (*M. javanica*). These three groups from different mammalian orders have varying dispersal capabilities and ecological requirements. Colugos glide through forest canopies, feed on young leaves, and are strictly arboreal which restricts them to forested habitat (Lim 2007). The Sunda pangolin spends much of its life in trees and is semi-arboreal, making dens in or around trees. Pangolins primarily feed on ants and termites using their powerful claws for digging through soil (Lim & Ng 2008). Mouse deer

include the smallest ungulates on the planet, with pencil thin legs and weighing only 2-8 kg, depending on the species. Mouse deer are terrestrial even-toed ungulates that feed on fallen fruits, shoots, seeds, and stems, and prefer to live in thick brush near swampy areas in tropical forests or mangroves (Prothero & Foss 2007). These arboreal, semi-arboreal, and terrestrial lifestyles, and varying food requirements, of these three genera could have had profound and different influences on the timing and modes of evolutionary diversification within each group. We consider these ecological differences when comparing colugo, mouse deer, and pangolin evolutionary histories.

Colugos are currently classified as two species in monotypic genera: the Sunda colugo, *Galeopterus variegatus*, and the Philippine colugo, *Cynocephalus volans* (Wilson & Reeder 2005). Recently however, molecular genetic data indicated that *Galeopterus* would be more accurately divided into six or seven species and *Cynocephalus* into two species (Fig. 4.1) (Mason et al. submitted). Substantial levels of cryptic genetic diversity among colugos is highlighted by maintained genetic isolation and strong geographic sorting of individuals into populations that are diverged on a scale of millions of years, with little evidence for recent (Pleistocene) genetic exchange between them. The Sunda colugo is distributed across Sundaland, a composite of Gondwanan continental fragments that is the product of dynamic geologic events that have created a broad shelf with shallow seas separating the major islands of Sumatra, Borneo, and Java (Metcalfe 2011). The Philippine colugo is found only in the southern Philippine islands that comprised the Pleistocene super-island of Greater Mindanao.

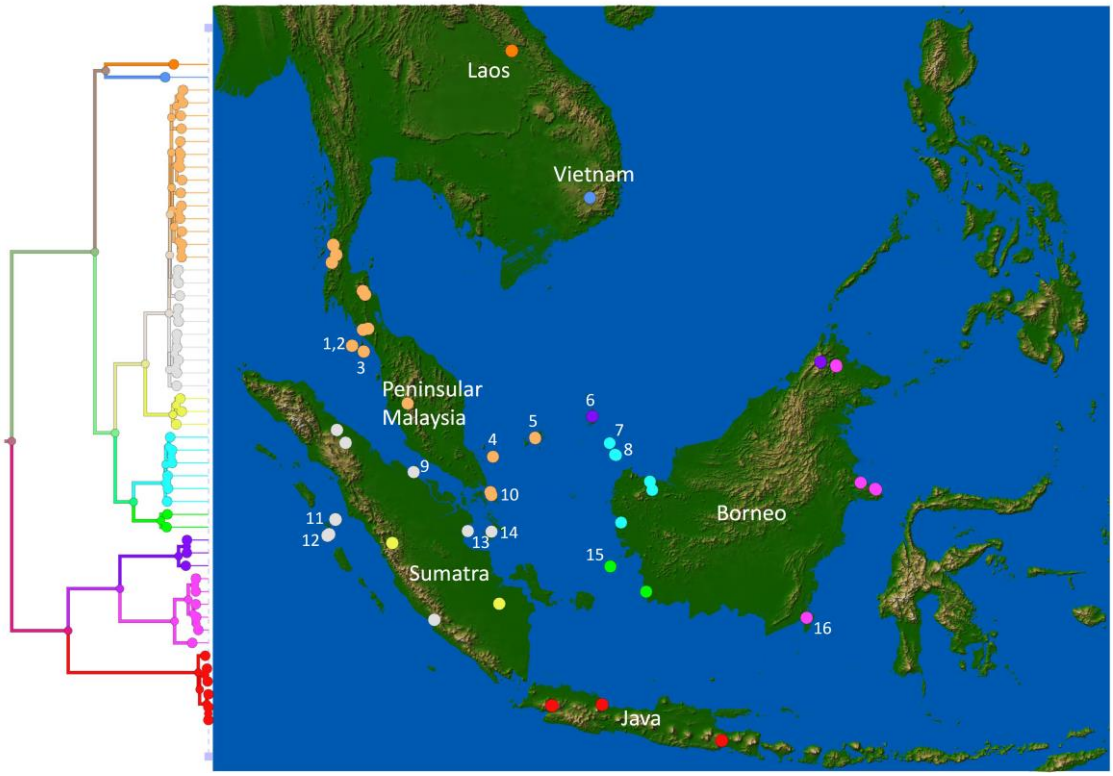


Figure 4.1. Colugo mitochondrial maximum likelihood phylogeny where each color is a supported monophyletic clade or highly divergent lineage and map of Sundaland. Geographci locations are color coded by monophyletic group and dots on the map show sampling collection locations of Sundaic colugos. Geographic labels of major islands are labeled on the map and minor islands are labeled with numbers: 1) Ko Rawi 2) Ko Adang 3) Pulau Langkawi 4) Pulau Aur 5) Pulau Siantan 6) Pulau Bunguran 7) Pulau Subi-besar 8) Pulau Serasan 9) Pulau Rupat 10) Pulau Bintan 11) Pulau Pini 12) Pulau Tanahbala 13) Pulau Bakong 14) Pulau Penuba 15) Pulau Karimata 16) Pulau Sebuko.

Colugos are an obligately arboreal species that represent an ideal taxon for inferring past forest distributions. They are broadly distributed across more than 50 Sundaic islands, and have restricted dispersal capabilities that result from their specialized adaptations for gliding through high-density forest canopies. Colugos glide with the use of their patagium, the most extensive gliding membrane of any known vertebrate. The patagium maximizes its surface area by interconnecting all appendages

and digits, and non-retractable claws aid in hanging from branches but hinder terrestrial movement (Beard 1993, Lim 2007). While well adapted for gliding, the patagium restricts colugos terrestrial movement to a series of short hops when on the ground, and even when climbing up tree trunks (Lim 2007). Thus, colugos' ability to move outside of forested habitat is very limited, restricting them to forested corridors for dispersal. Moreover, long-distance overwater dispersals are unlikely as colugos can glide, but can not achieve true powered flight. We hypothesized that deciphering colugo evolutionary history and diversification events could provide unique inferences regarding when and where forested connections were present in the past between different geographical regions within Sundaland.

Islands across the Sunda shelf are currently isolated by large expanses of shallow seas; however contemporary island locations and structure, and exposed land-area are poor representations of the dynamic geologic, climactic, and environmental histories of the region. Many major geological changes in the Miocene (23 Mya -5.3 Mya) impacted climate and biotic diversification in Sundaland. The collision of the Australian and Asian plates (~23 Mya) closed the deep ocean valley between them, which altered ocean currents and brought warmer water to flood the region, promoted uplift of the Sunda shelf, and increased the area of the surrounding shallow-seas (Hall 2013). The collision between Australian and Asian plates and subduction of the Dangerous Grounds under northern Borneo (~10 Mya) resulted in the formation of the Bornean highlands, which likely contained snow-capped mountains at that time (Hope 2004, Hall 2013). Throughout the Miocene the Pacific Ocean gradually flooded the previously exposed

Sunda shelf through rift basins south of Vietnam and in the Gulf of Thailand eventually resulting in isolation of mainland from Borneo ~5 Mya (Shoup et al. 2013, Hall 2013). Subduction of the Celebes Sea under southern Sabah and Sulu arc formed a chain of volcanic islands from NE Borneo to the Southern Philippines (~10 Mya) (Hall 2013). By the early Pliocene (~5 Mya) much of the modern geologic structure of the Sunda shelf was completed. The early Pliocene (5.3 Mya - 3.6 Mya) was characterized by elevated sea levels well above present day levels (Uba et al. 2007), followed by major fluctuations in climate begining in the late Pliocene (~3.2 - 2.6 Mya) which continued throughout the Quaternary (2.6 Mya - Present). During this time ~50 glacial cycles caused frequent and periodic oscillations of sea levels throughout Sundaland (Woodruff 2010). These glacial cycles were characterized by long glacial periods and relatively short interglacial periods.

Low sea levels during glacial periods exposed large expanses of the Sunda shelf, opening potential overland dispersal corridors for terrestrial taxa that were subsequently closed during the following interglacial period. These fluctuating sea levels likely had a major impact on biotic diversification patterns by repeatedly interconnecting previously isolated landmasses and then isolating them again. Sea level fluctuations within the last one million year resulted in three major environmental scenarios: 1) very low sea levels (-120m) indicative of glacial maxima, and high land-area exposure were present ~37% of the last million years; 2) intermediate sea levels (-40m-50m), were present ~55% of the last million years, resulting in about half of the Sunda shelf being exposed with land bridges between Borneo, Peninsular Malaysia, and Sumatra; and 3) interglacial periods

with high sea levels (-0m), characterized by modern isolation of islands and low land-area exposure representing refugial forest and biotic distributions comprised only ~8% of the last million years (Cannon et al. 2009, Wilcove 2011, de Bruyn et al. 2014). Overland connections were present during scenarios 1 and 2, however important ecological differences distinguish them. The high land-area exposure and dearth of inland bodies of water expected under scenario 1 would have resulted in weaker seasonal monsoons and overall drier climate that might have hindered forest growth, while scenario 2 would have provided greater moisture for the exposed Sunda shelf (Hope 2007, Woodruff 2010, Lohmann et al. 2011). A drier climate during glacial maxima, together with the coarse sandy soils of central Sundaland (Slik et al. 2011), has led many to speculate that a savanna ran north-south from eastern Peninsular Malaysia to Java during glacial maxima (scenario 1) and specifically during the last glacial maxima (LGM) (~0.0015 Mya) (Heaney 1991, Bird et al. 2005). This corridor would have prevented dispersal of forest-dependent taxa between major landmasses like Sumatra and Borneo. However, variation in environmental conditions between glacial maxima could have provided conditions suitable for continuous forested connections to span east-west connecting Borneo with Sumatra and/or Peninsular Malaysia. Evidence for continuous forested connections has been indicated through spatially explicit modeling of geography, paleoclimactic, and geologic variables, as well as Dipterocarp species distribution models (Cannon et al. 2009, Raes et al. 2014, de Bruyn et al. 2014). Here we: 1) examine patterns of genetic differentiation and use ancestral area probability calculations to suggest how and when colugos arrived at their present locations; 2)

consider the probable time, extent, and location of forested corridors in Sundaland throughout the last 10 million years as evidenced by colugo evolutionary history; and 3) use comparisons of colugo phylogenetic history to that of new molecular data collected from pangolins and mouse deer, and published data amassed from studies of other SE Asian mammals and birds to propose how different modes of dispersal in mammals have influenced the timing and patterns of their diversification.

4.2 Materials and Methods

4.2.1 Museum Samples, Illumina Library Preparation, and Capture Hybridization

We sampled ~5mg of dried adherent tissue from colugo, mouse deer, and pangonlin museum specimens (Table 4.1). Although we sampled skin, muscle, cartilage, and bone, we preferentially collected dried adherent tissue from inside the more protected skull and nasal turbinate systems (brain and nasal crusties). DNA was extracted from museum tissues using ethanol precipitation procedures with large digestion volumes that included the carrier molecule linear acrylamide and excess proteinase K as detailed in Mason et al. (submitted) (Chapter III). All DNA was created into Illumina libraries and sequenced on the Illumina HiSeq 2000 or 2500 sequencing platforms. Enrichment for orthologous nuclear DNA sequences was done with DNA capture hybridization procedures that have been described previously (Mason et al. 2011, Mason et al. submitted). Primers for probe design, targeted sequences, and capture methodology were published in Mason et al. (submitted).

Table 4.1. Museum specimen information for *Manis* and *Tragulus* genera.

NMNH Cat. No.	Sex	Species	Date Collected	Latitude	Longitude	County	Island	Country
104598	M	*Manis javanicus*	6/24/1900	3.963680°	108.187035°		Natuna Islands: Bunguran	Indonesia
142460 / A49875	M	*Manis javanicus*	7/19/1905	-0.020507°	109.349547°	Pontianak	Borneo	Indonesia
317198	M	*Manis javanicus*	9/28/1960	5.943883°	116.685925°	Sabah, Ranau	Borneo	Malaysia
260592	F	*Manis javanicus*	11/8/1931	14.144818°	102.364671°	Ban Ku Khano		Thailand
356430	M	*Manis javanicus*	7/23/1966	12.288028°	108.219792°	Mount Santra Area		Vietnam
268293	F	*Tragulus javanacus*	3/30/1908	-7.723297°	108.490079°	Pangandaran: Dirk Vries Bay	Java	Indonesia
267202	F	*Tragulus williamsoni*	4/4/1936	19.248491°	100.310616°	Huai oi		Thailand
144321	M	*Tragulus kanchil*	1/24/1907	0.900587°	102.669766°		Sumatra: Pulo Tebing Tinggi	Indonesia
153748	M	*Tragulus kanchil*	8/8/1908	-2.544825°	110.204819°	Kendawangan River, Lanchut	Borneo	Indonesia
143491	M	*Tragulus kanchil*	1/2/1906	4.143283°	98.157660°	Aru Bay	Sumatra	Indonesia
019191 / A34912	M	*Tragulus kanchil*	12/7/1887	5.438124°	118.106412°	Kinabatangan River	Borneo	Indonesia
104608	F	*Tragulus napu*	7/9/1900	3.931526°	108.174188°		Natuna Islands: Bunguran	Indonesia
144135	F	*Tragulus napu*	12/18/1906	0.795817°	101.799352°	Siak River	Sumatra	Indonesia

Table 4.1 Continued.

NMNH Cat. No.	Sex	Species	Date Collected	Latitude	Longitude	County	Island	Country
141175	F	*Tragulus napu*	3/13/1905	1.082996°	97.795960°	Pulo Nias, Mojeia River	Sumatra	Indonesia
124928	M	*Tragulus napu*	9/20/1904	-2.903702°	107.548895°	Tanjong Batu	Billiton	Indonesia
104617	M	*Tragulus napu*	6/9/1900	2.512672°	109.045964°		Natuna Islands: Sirhassen (Serasan)	Indonesia
115506	F	*Tragulus napu*	9/1/1902	0.767655°	103.708186°		Rhio Archipelago: Pulo Sugibawa	Indonesia
105000	M	*Tragulus napu*	10/13/1900	2.789633°	104.169407°	Pahang	Pulo Tioman	Malaysia
144423	M	*Tragulus napu*	3/28/1907	0.933282°	104.067187°		Rhio Archipelago: Pulo Setoko	Indonesia
104416	M	*Tragulus napu*	12/8/1899	6.357468°	99.810267°	Kedah	Pulo Langkawi	Malaysia
153753	M	*Tragulus napu*	8/16/1908	-2.544825°	110.204819°	Kendawangan River	Borneo	Indonesia
NCBI Reference Sequences:								
NC_016008		*Manis pentadactyla*						
NC_020753		*Tragulus kanchil*		18.897182°	103.882528°			

4.2.2 Sequence Assembly and Analysis

Colugo, mouse deer, and pangolin mitochondrial genome (mtDNA) sequences were assembled with SOAP-denovo2, or via reference assembly (bwa) (Li & Durbin, Luo 2012). Some colugo mitochondrial genomes were assembled from off-target nuclear capture sequences (Mason et al. submitted). Mitochondrial genomes were aligned with Mafft v7.127, and refined by manual adjustment in Geneious R7 (Kearse 2012, Katoh 2013).

Maximum likelihood phylogenies were generated for each alignment using RAxML v8.1.17, with 1000 bootstrap iterations (Stamatakas 2014). The biparental dataset (115 kb) was composed of nearly-neutral X-chromosome loci (81 kb) and autosomal coding loci (34 kb). Proposed species groups for the Sunda colugo were based on phylogenetic data detailed in Mason et al. (submitted). Maximum composite likelihood genetic distance calculations (gamma = 4) were calculated in MEGA v6.06 (Tamura 2013). Mitochondrial DNA genome sequences from eighteen mouse deer individuals were included in phylogenetic analyses if the consensus sequences covered >75% of the genome (16,358 bp following removal of poorly-aligned sites in the alignment). Mitochondrial DNA genome sequences from seven pangolin individuals were included in phylogenetic analyses if the consensus sequences covered >30% mtDNA genome base coverage were used in phylogenies. The pangolin phylogeny in figure 4.2 was restricted to only bases covered by the E. Bornean pangolin MJA317198 to limit effects of missing data (4644 bps).

4.2.3 Divergence Time Calculation

Due to the absence of relevant colugo fossils, divergence times were calculated for these taxa based on secondary molecular calibrations derived from whole genome one-to-one orthologous protein coding DNA sequences for 12 mammalian taxa (Mason et al. submitted). Sequences for nine taxa were derived from OrthoMaM v9 (Douzery et al. 2014). Orthologous colugo and treeshrew sequences were added to the nine-taxon data matrix as detailed in Mason et al. submitted (Chapter III). Seven mammalian fossil calibrations were used to constrain molecular estimates of *Galeopterus/Cynocephalus* divergence time (Mason et al. submitted). The upper and lower bounds of the 95% confidence interval for the divergence time between *Galeopterus* and *Cynocephalus* were used to calibrate all subsequent colugo specific time-trees derived from capture-based hybridization or low coverage sequencing.

Divergence times within the three mouse deer species (*T. javanica*, *T. napu*, and *T. kanchil*), were similarly calibrated based on the 95% confidence intervals of a secondarily derived divergence time (15.9-29.1 Mya) between Sundaic mouse deer and the African water chevrotain *Hyemoschus aquaticus* (accession number: JN632650) (Hassanin et al. 2011). The African pangolin *Manis tetradactyla* (*M. tetradactyla*) was used as the outgroup and calibrative branch for calculating divergence dates for pangolins. The 95% confidence interval of the divergence time (16.9 – 35.7 Mya) between *M. tetradactyla* and the species *M. javanica* and *M. pentadactyla* based on the molecular estimates described in Meredith et al. (2011) was used to calibrate *Manis* phylogenies. We estimated divergence times using MCMCtree v4.8a (Dos Reis et al.

2014), assuming an autocorrelated rates model with a soft calibrations within each group of colugos, mouse deer, and pangolins. Calculations were performed with exact likelihood and an HKY-85+Γ model of sequence evolution. Datasets were reduced to one taxon per divergent lineage, where individuals were selected by having the greatest percent genome coverage and sequence depth as reported in Mason et al. submitted. The rate parameter β for the rgene_gamma prior was estimated by calculating the clock like substitution rate for the entire phylogeny in baseml v4.8a using a nucleotide alignment, and rooted phylogeny with branch lengths and all available fossil calibration point estimates. We chose a diffuse gamma distribution shape parameter $\alpha = 1$ for all MCMCtree calculations to emulate the uncertainty in fossil calibrations. When $\alpha = 1$ the mean (m) equals the standard deviation (s) of a gamma distribution. The clock like substitution rate was used as the mean of the gamma distribution, and since $\alpha = 1$, $m = s$. We calculated $\beta = m/s^2$.

4.2.4 Ancestral Area Probability Analyses

Ancestral area probabilities were calculated for mtDNA and biparental data sets with BioGeoBears (phylo.wikidot.com/biogeobears, Matzke 2013, Matzke 2014). *Tragulus kanchil* was divided into three separate operational taxonomic units (OTUs) for ancestral area probability calculations to represent the diversity within the species. Settings for max_range_size were five for mouse deer and seven for colugos. These settings allow all seven OTUs of colugo and all five OTUs of mouse deer to be represented under one ancestral range if appropriate. Six different models were run DEC, DEC+J, DIVA, DIVA+J, BAYAREA, and BAYAREA+J. Where DEC is

dispersal extinction cladogenesis (i.e. Lagrange) (Ree & Smith 2008), DIVA is dispersal-vicariance analysis (Ronquist 1997), BayArea (bayesian), and +J adds parameter jump-dispersal. Corrected Akaike information criterion (AICc) values were compared as a measure of model fit to data between non-nested models where the lowest AICc indicated best fit. Significant differences between nested models were detected with likelihood ratio tests (i.e. DIVA is nested within DIVA+J).

Comparative biogeographic inferences were derived by comparing different evolutionary histories between species, comparing times of divergence to known geologic and ecological events, and considering species specific ecological requirements. A cytochrome-b (cytb) phylogeny was constructed that incorporated additional sequences from Endo et al. (2004), to increase geographic and taxonomic sampling, and validate sequences within *Tragulus*. A COI phylogeny was also constructed to validate pangolin sequences with additional sequences from NCBI (fig A4.3) (Zhang 2015).

4.3 Results

4.3.1 Colugo Phylogeography

Colugo phylogenies for maternal, biparental, and paternal evolutionary histories showed significant structure and deep divergence times within each described species of colugo (Figs. 4.2-4.3) (Mason et al. submitted). Phylogenies show evidence for divergent, monophyletic evolutionary lineages from Laos, Vietnam, Peninsular Malaysia+Sumatra, Java, W. Borneo, and E. Borneo (Fig. 4.1, Fig. 4.2). Each of these seven populations represent potential species groups that diverged earlier than the late Pliocene (>3.6 Mya) (Fig. 4.3) (Mason et al. submitted).

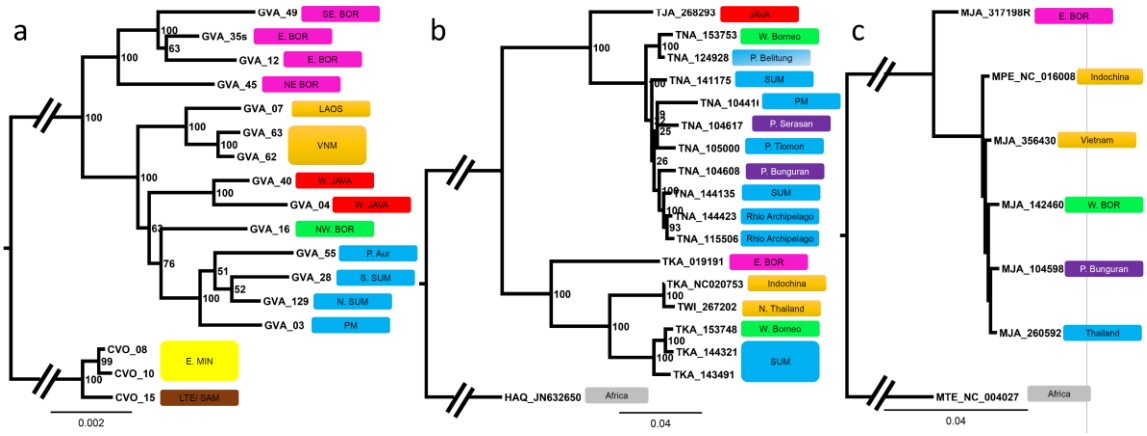


Figure 4.2. Maximum likelihood phylogenies constructed from A) Sunda colugo (GVA) and Philippine colugo (CVO) biparental sequence-based phylogenies, B) lesser mouse deer (TKA) mtDNA, greater mouse deer (TNA) mtDNA, Javan mouse deer (TJA) mtDNA, and C) Sunda pangolin (MJA) mtDNA. Individuals are labeled with their sample ID, museum ID, or NCBI accession number. Individuals are labeled with color-coded geographic locations that represent the sample collection location. Bootstrap values are based upon 1000 replicates.

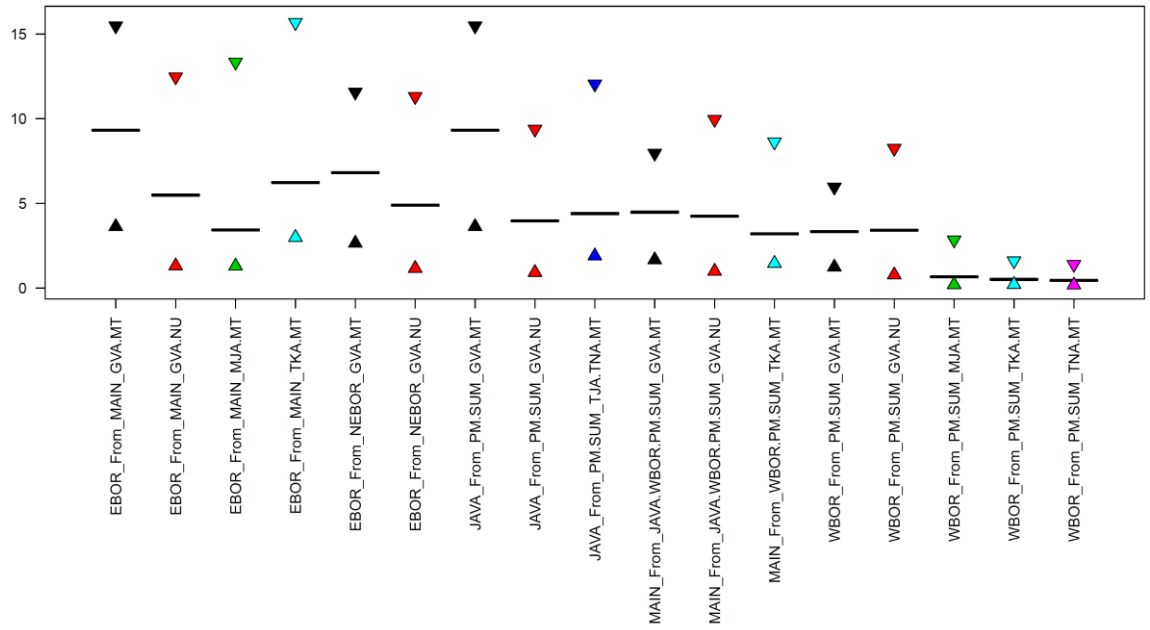


Figure 4.3. Divergence times and 95% confidence intervals of major phylogeographic splits based on colugo mtDNA alignments (GVA.MT), colugo nuclear, biparental alignments, lesser mouse deer mtDNA alignments (TKA.MT), greater mouse deer mtDNA alignments (TNA.MT), and pangolin mtDNA alignments (MJA.MT).

4.3.2 Mouse Deer and Pangolin Phylogeography and Divergence Times

Mouse deer mitogenome-based phylogenies support monophyly of *Tragulus napu* and *T. kanchil*, and identify *T. javanicus* as a divergent sister-taxon to *T. napu* (Fig. 4.2). Substantial phylogenetic structure sorts individuals of *T. kanchil* by geographic location into three deeply divergent lineages: East Borneo, Indochina, and Peninsular Malaysia, confirming previously reported mtDNA genetic divergence between these geographically isolated populations (Fig. 4.4) (Endo et al. 2004). By contrast, *T. napu* shows less geographic sorting of individuals and reduced mean intraspecific pairwise genetic divergence (avg.=1.6%) relative to *T. kanchil* (avg.=4.7%) (Fig. 4.4) (Endo et al. 2004). The one representative of *T. williamsoni* was very similar to the Indochinese *T.*

kanchil populations (0.5% mtDNA genetic divergence). The divergence between *T. kanchil* populations from E. Borneo and other *T. kanchil* populations was large (~8%) and exceeded divergence estimates between the two described species, *T. javanica* and *T. napu*.

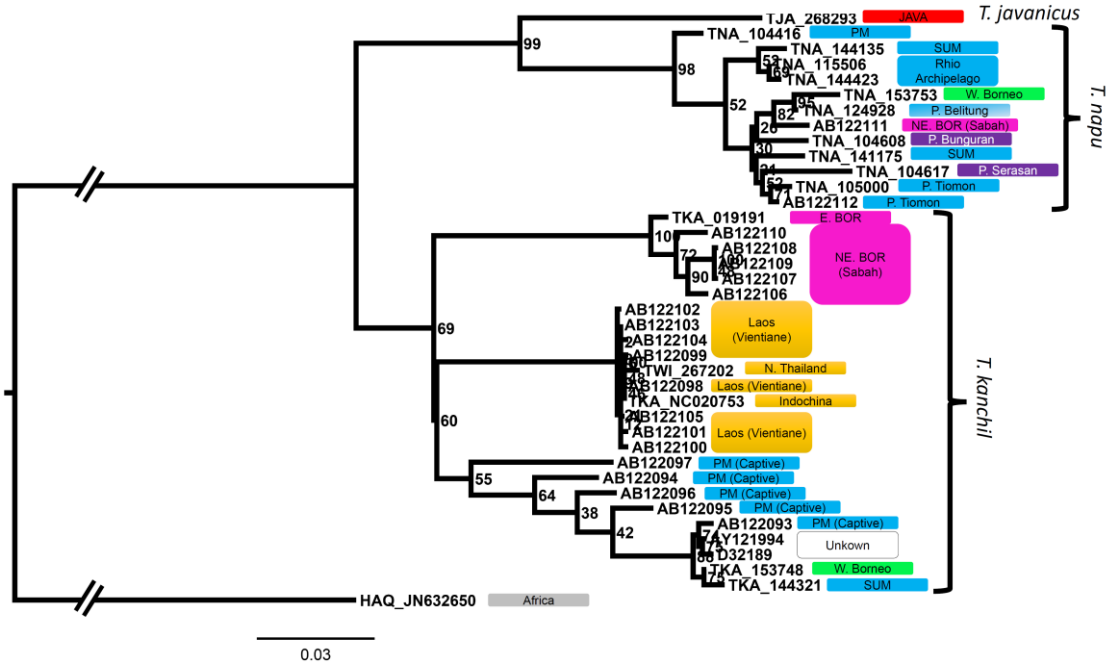


Figure 4.4. *Tragulus* cytb maximum likelihood phylogeny including data from Endo (2004), and outgroup *Hyemoschus aquaticus* (HAQ). Individuals are labeled with a species abbreviation followed by their museum accession number, or NCBI accession number for individuals from Endo et al. (2004). Labels are followed by color coded geographic labels indicating sample collection location. Bootstrap values were derived from 1000 bootstrap replicates. Cytb sequences from captive individuals of ambiguous origin are noted on the phylogeny

By comparison pangolins show little mtDNA phylogenetic structure between Indochina, Peninsular Malaysia, and West Borneo (Fig. 4.2). However, we observed

significant genetic differentiation between pangolins from E. Borneo and more western Sundaic regions (Figs. 4.2-4.3).

4.3.3 Ancestral Area Probabilities

Borneo was estimated to be the most probable area of origin for the Sunda colugo based on all biogeographic models regardless of genetic marker (Fig 4.5, fig A4.1). Analysis of the mouse deer phylogeny and timescale also supports Borneo as the most probable ancestral area of origin based on the BAYAREA and BAYAREA+J models. Borneo was the most probable single geographical origin for *T. javanica, T. napu*, and *T. kanchil* based on the DIVA and DIVA+J models f (fig A4.2). The most probable area of origin for mouse deer spanned across Peninsular Malaysia + Sumatra, Java, and Borneo together under DIVA, DIVA+J, DEC and DEC+J models. This indicates only that mouse deer were Sundaic in origin.

The best fitting ancestral area model was DIVA+J for colugo mtDNA and biparental time-trees with AICc values of 29.7, and 30.2 respectively. DIVA was the best fitting model for the mouse deer time-trees with AICc of 32.6. Adding jump-dispersal (J) significantly improved the fit of the model to the colugo dataset, but no significant improvement was identified for the mouse deer dataset.

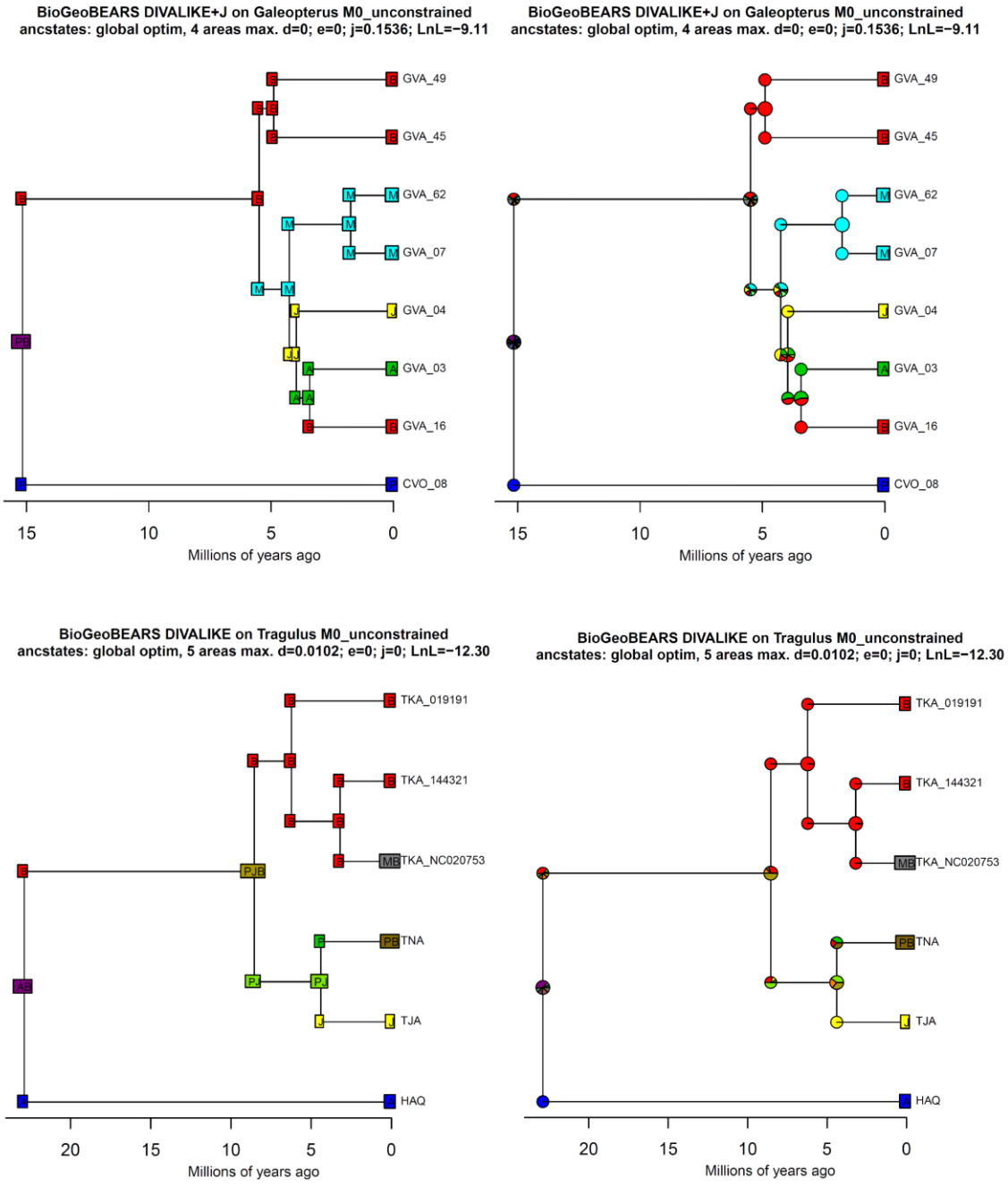


Figure 4.5. Ancestral area probability calculations using the all nuclear colugo time-tree (top) and full mtDNA genome time-tree for mouse deer (bottom) under the best fitting models of DIVA+J and DIVA respectively. Trees on the left show the most probable ancestral area for colugos or mouse deer while trees on the right show color-coded pie-charts indicating the probablility of each geographic region or combination of geographic regions.

4.4 Discussion

Sundaland is a region well known for high densities of biological diversity. Borneo and mainland Southeast Asia have been described as areas of refuge for biological diversity throughout the dramatic geologic and climatic changes that have taken place since the early Miocene (de Bruyn et al. 2014). Here we used phylogenetic methods, divergence dating, ancestral area reconstructions, ecological information, and the geological and climactic histories of Sundaland to develop a hypothesis for the sequence of events that have resulted in the present day distributions of colugos, mouse deer, and pangolins. We also use these results to infer probable forested connections between currently isolated landmasses at different times in the past.

4.4.1 Colugo Origins and the 'Out of Borneo' Hypothesis

Ancestral area reconstructions support Borneo as the most likely origin for Sunda colugos (Fig. 4.5, fig A4.1). This is certainly plausible because much of western and northern Borneo was subaerial throughout the Cenozoic (Moss and Wilson 1998), and phylogenetic diversity within Borneo is much greater and older than that of other Sundaic and Indochinese populations. However, the origin of all extant colugos remains unresolved by our ancestral area reconstructions (Fig 4.5, fig A4.1). Nevertheless, we hypothesize that Borneo might have been the origin of both extant colugo genera, because much of the southern Philippines (Greater Mindanao) were submerged, or distantly isolated hundreds of kilometers to the southeast from the northern Philippine islands, or recently formed through volcanic activity (Hall et al. 1995, Hall 2002, Steppan et al. 2003, Hall 2013). The lowest observed sea levels in the Cenozoic prior to

the Pliocene occured ~10.5 Mya (Meijaard 2004). Interestingly this date approximately coincides with the estimated divergence time between to the colugo genera at 11.3 Mya. In addition, the chain of volcanic islands of the Sulu archipelago connecting NE. Borneo and the Philippines were formed about ~10 Mya (Hall 2013). This corroborative evidence provides a plausible mechanism for the dispersal of colugos from Borneo to the Philippines across the Sulu archipelago during low-sea stands. This 'out of Borneo' hypothesis states that colugos originated within Borneo and subsequently migrated to their present day distributions in the southern Philippines, the Indochinese Peninsula, and Java.

The ancestor of present day E. Bornean colugo populations was initially isolated ~9 Mya, while nuclear gene flow continued between all Sundaic colugos until ~5 Mya, when nuclear genetic exchange ceased (Fig. 4.2). The cessation of nuclear gene flow indicates that genetic isolation was established and subsequently maintained between two colugo populations that likely resided on the Indochinese Peninsula and Borneo ~5 Mya. The initial mtDNA isolation may have been influenced by maturation of Bornean highlands and river systems into their more modern representation by ~8-10 Mya (Hall 2013). Nuclear genetic isolation of modern E. Bornean colugos from the rest of the Sundaic colugos (in PM and Indochina) was likely facilitated by flooding of the Sunda shelf that closed the previously continuous land connection between mainland and Borneo ~5 Mya (Shoup 2013). These high sea-levels were maintained for 1.5 Mya (Uba et al. 2007) and would likely have provided sufficient time for genetic incompatibilities to accumulate between the two colugo populations (Hedges et al. 2015).

An early isolation within Borneo was likely the cause for the significant genetic differentiation in colugo maternal and biparental histories observed between modern northeastern (Sabah) and eastern (Sabah & E. Kalimantan) Bornean colugos, despite the lack of dispersal barriers (except rivers) between these regions. This pattern of isolation is similar to that proposed to explain subspecies differences between bird populations from Sabah and E. Kalimantan (Sheldon et al. 2009).

Eustatic sea-level fluctuations would have caused repeated overland connections between Mainland/Peninsular Malaysia and other areas of the Sunda shelf from the late Pliocene onward. These connections seem to have been utilized in a series of rapid dispersal events leading to explosive speciation. This is indicated by the unresolved starlike phylogeny of colugos evolving from an ancestral Southeast Asian mainland population out into the Sunda shelf to create modern lineages of Java, W. Borneo, and Peninsular Malaysia + Sumatra, as identified in the nuclear DNA trees (Fig. 4.2). The current W. Bornean lineage likely dispersed from the mainland and colonized unoccupied habitat, or alternatively colonized and replaced ancestral Bornean populations (members of the current E. Bornean clade) if they were present. Colugos dispersing from the ancestral mainland population also likely colonized Sumatra by either taking over vacant habitat or replacing currently existing colugo populations. However, gene flow between Sumatran and Peninsular Malaysian colugos continued, when sea levels were less than ~30m below present levels well into the Pleistocene

4.4.2 Java

The origin of the current Javan colugo populations appears more complex, as indicated by topological and divergence time differences between the maternal and biparental phylogenies (Figs. 4.2-4.3). While all modern Javan colugos have retained an ancestral E. Bornean-like mtDNA genome, their nuclear history appears to be more recently derived from the ancestral mainland population, which radiated throughout the western Sunda shelf in the Pliocene. Most reconstructions indicate that the majority of Java was submerged until as recently as ~5 Mya (Hall 2013). If colugos were present at this time they would have been restricted to mountainous refugia in W. Java. We hypothesize that a 'local' population of ancestral Bornean colugos was present on Java and that male-biased dispersal of colugos from the ancestral mainland population resulted in mitochondrial capture of an ancestral Bornean mtDNA genome by the colonizing ancestral mainland individuals (Petit & Excoffier 2009). This mitochondrial replacement could have been facilitated by at least two factors: 1) Haldane's Rule resulting in sex-biased fitness loss of hybrid males, and 2) generally male-biased dispersal patterns in mammals (Petit & Excoffier 2009). Haldane's rule suggests that if a fitness loss is observed in hybrid offspring of two divergent populations (through secondary contact) the fitness loss will occur in the heterogametic sex (Haldane 1922). A greater relative fitness loss in hybrid males than in hybrid females would have promoted mitochondrial capture of ancestral E. Bornean-like mtDNA if hybrids were formed between dispersing males from ancestral mainland populations and females within Javan populations with Eastern Bornean ancestry. In addition, Petit & Excoffier (2009) have

shown that molecules of inheritance that undergo less intraspecific gene-flow (i.e., mitochondrial DNA) in colonizing 'wavefronts' are more susceptible to genetic introgression. The vast majority of mammals exhibit male-biased dispersal patterns and show evidence for mitochondrial introgression from local populations to the male-biased colonizing populations (Petit & Excoffier 2009, Toews and Brelsford 2012).

4.4.3 Comparative Biogeography

Some observed phylogenetic differentiation patterns in colugos, mouse deer, and pangolins are different than most Sundaic mammals studied to date. Molecular dating analyses report late Miocene (mtDNA) and early Pliocene (nuclear) divergence times between East and West Borneo for the colugo. This pattern of deep diversification across Borneo has rarely been reported in mammals (Han 2000), but is becoming increasingly apparent from avian biogeographic analyses (Molye 2005, Sheldon 2009, Sheldon 2015). For example, mtDNA evidence of reciprocal monophyly from white-crowned forktails *Enicurus leschenaulti* (*E. l.*) suggests that montane environments in northern Borneo have created a dispersal barrier between northeast (Sabah) and west Bornean (Sarawak) lowland *E. l. frontalis* subspecies (Moyle 2005). The mountainous environment in north central Borneo also harbors a distinct subspecies of white-crowned forktail (*E. l. borneensis*) suggesting that it limited east west dispersal across Borneo. In addition, phylogenies of the oriental magpie-robin *Copsychus saularis* (*C. s.*) depict similar patterns of subspecies differentiation between northeast (*C. s. musicus*) and west Borneo (*C. s. adamsi*) with an estimated ~1.2 Mya mtDNA divergence between subspecies (Sheldon et al. 2009). The mountain black-eye *Chlorocharis emiliae* (*C.*

emiliae) also shows complete northeast (Sabah) isolation from west (Sarawak) individuals (Gawin 2014). Despite the differences in vagility between volant avian species versus gliding colugos, it is remarkable that some avian species possess subspecific differentiation that corroborates the species-level differentiation we observe among Bornean colugos.

We also observe a strong east west genetic differentiation within Borneo among lesser mouse deer (*T. kanchil*) (Fig. 4.2, Fig 4.4). This pattern is not seen in the greater mouse deer (*T. napu*) (Fig. 4.4). We hypothesized that all mouse deer species would have less genetic structure and more vagility than both Sundaic colugos and pangolins because mouse deer are fully terrestrial and possibly better suited to disperse across alternative environments, while colugos are strictly arboreal and the Sundaic pangolins are semi-arboreal. Surprisingly, the lesser mouse deer mimics the deep divergence times and highly structured evolutionary history of Sundaic colugos, while greater mouse deer show little geographic sorting of individuals across their broad range from Peninsula Malaysia to E. Borneo (Fig 4.2). Lesser mouse deer exhibit threefold greater interspecific mtDNA divergence (4.7%) between populations than the greater mouse deer (1.6%). Perhaps the larger body size of the greater mouse deer facilitates dispersal across larger distances and more variable ecotypes, like montane environments in Borneo, whereas the lesser mouse deer would be more restricted in range and dependent on evergreen forested ecotypes.

Sundaic pangolins have relatively high genetic divergence values between east and west Borneo similar to that seen in colugos and the lesser mouse deer. However,

Peninsular Malaysian and West Bornean pangolins recently diverged in the middle Pleistocene, much like greater and lesser mouse deer, but unlike colugos which diverged much earlier in the late Pliocene. Mitochondrial data from forest dependent mammalian and avian taxa on Peninsular Malaysia and Sumatra are more often genetically similar than Bornean taxa which represent a more divergent sister clade (Leonard et al. 2015, Sheldon et al. 2015). This is supported in colugo, mouse deer and pangolin phylogenies, however colugos on Peninsular Malaysia/Sumatra and W. Borneo diverged much earlier, in the Late Pliocene (mtDNA 3.3 Mya, biparental 3.4 Mya), when compared to the relatively recent Middle Pleistocene divergence times of the lesser mouse deer (0.51 Mya), greater mouse deer (0.45Mya), and Sundaic pangolins (0.61 Mya). This fits well with the 1.31 average divergence time between Peninsular Malaysia and Sumatran avian and mammalian populations, with a maximum divergence time of 3.9 Mya among chestnut-winged babblers, *Stachyris erythroptera* (Leonard et al. 2015). This suggests that forested corridors were present and facilitated gene flow between forest-dependent taxa from Peninsular Malaysia/Sumatra and W. Borneo during more recent Pleistocene periods. By contrast, colugos show no evidence for a similar Pleistocene exchange.

The mountains separating Sabah from Sarawak significantly limited east west interactions across northern Borneo for several bird and avian taxa: 1) *E. l. frontalis*, 2) *C. s. musicus* and *C. s. adamsi*, 3) *C. emiliae*, 4) *G. variegatus*, 5) *T. kanchil*, 6) *M. javanica*. For colugos this could very well be due to mountain formation and river maturation in northern and central Borneo (Hall 2013). However, these barriers are not impermeable; one sample of west Bornean *C. s. musicus* is present within the NE.

Bornean *C. s. adamsi* distribution (Sheldon et al. 2009). Similarly mtDNA evidence of one colugo specimen from Bunguran in the Natuna island chain recently diverged from the NE. Bornean population of colugos (Fig. 4.2). This is the only evidence for long distance migration within the late Pleistocene observed in our colugo sampling, which might have been facilitated by the predicted large expansions of montane and lowland rainforest during glacial maxima (Cannon et al. 2009). This could have allowed NE Borneo to be connected to W. Borneo, however we propose that the northern mountains of Borneo were still a substantial barrier for dispersal at that time. The strong separation of eastern and western gene pools indicates this genetic isolation has been largely maintained. However, further sampling of colugos across N Borneo is required to determine whether genetic exchange occurs or has occurred between populations. We think genetic exchange is unlikely as West and East Bornean colugo populations diverged ~5.5 Mya (Fig. 4.3) which is more than twice the average time to speciation (~2 My) among plants and animals (Hedges et al. 2015).

This is contrasted by little prior evidence for genetic differentiation between east and west Bornean populations of other mammals. The Bornean orangutan (*Pongo pygmaeus*), for example shows no geographic structure and has very recent (0.176 Mya) mtDNA coalescent times for populations dispersed across Borneo (Arora et al. 2010). Phylogenetic analyses from a limited number of samples of *Macaca fascicularis* from Sabah, Sarawak, and Kalimantan Borneo also show little evidence for significant genetic differentiation within Borneo (Tosi & Coke 2007). Similarly, there is no observed geographic structuring of phylogenies within Sundaic pigs (*Sus barbatus*) (Larson 2007,

Leonard 2015). One limiting factor is that the sampling schemes of many other mammalian studies in the region were biased for interisland comparisons and were thus incapable of detecting genetic differentiation within Borneo (Steppan et al. 2003, Ziegler et al. 2007, Patou et al. 2010, Den Tex 2010, Achmadi et al. 2013). We believe that additional sampling focused on forest-dependent or low vagility mammals within Borneo could reveal further evidence of east west genetic differentiation on the island, similar to our observations from colugos, mouse deer, and pangolins.

4.4.4 Recent Population Diversification within Satellite Islands of the Sunda Shelf

Colugo evolutionary histories show evidence for one example of recent large scale movement of genetic material, however many cases of recent Pleistocene exchanges were found in nuclear and mtDNA time-trees between colugos on large landmasses and surrounding satellite islands. During Pleistocene glaciation cycles small satellite islands were readily connected to neighboring large landmasses with small reductions in sea level (-15m-55m). This would have enabled frequent migration from the large neighboring landmasses of Peninsular Malaysia, Sumatra, and Borneo, and perhaps resulted in genetic replacement of island colugo populations. Islands in the Rhio Archipelago contain populations of colugos with low mtDNA genetic divergence (<1%) from Peninsular Malaysian colugos (Fig. 4.1). Colugos on islands of Pulau Pini and Pulau Tanahbala from the Batu islands, Pulau Bakong and Pulau Penuba from the Rhio Archipelago all diverged recently from Sumatran colugos (Fig. 4.1). Colugos from the lesser Natuna islands (Pulau Serasan and Pulau Subi-besar) are closely related to NW Borneo populations (Fig 4.1), while a colugo from Pulau Karimata was found to be very

similar to a SW Bornean colugo. A similar evolutionary pattern was found between red spiny rats (*Maxomys* surifer) on the Bornean mainland and red spiny rats on Pulau Karimata (Gorog et al. 2004). Also a colugo from Pulau Sebuku is most similar to other E. Bornean colugos (Fig. 4.1). Based on these patterns we can hypothesize that colugos from Pulau Tioman and Pulau Perhentian will be found phylogenetically similar to Peninsular Malaysia, and colugos from Pulau Banggi will be very similar to NE. or E. Bornean colugo populations.

Despite the low genetic divergence between Peninsular Malaysian and Sumatran colugos (~1%), populations from these major landmasses and surrounding islands are reciprocally monophyletic. This means that insular colugos are most similar to colugos on the closest major landmass, and that island colugo populations had multiple origins. Interestingly, satellite islands of Pulau Langkawi, Ko Rawi, Ko Adang, Ko Tarutao, Pulau Aur, Pulau Bakong, Pulau Tuangku, and Pulau Karimata all harbor dwarf colugo populations. This body size reduction adaptation was presumably an adaptation in response to limited island resources (Heaney 1978, Dunham 1978). Our mitochondrial analysis of colugos from Peninsular Malaysia, Sumatra, and neighboring islands show at least four independent origins of island dwarfism (Pulau Adang, Pulau Langkawi, Pulau Aur, Pulau Bakong) (fig A4.3). These convergent adaptations must have evolved more recently than the divergence of Peninsular Malaysian and Sumatran colugos, i.e. within the last 500,000 years (190,000 years – 970,000 years) (Fig. 4.3) (Mason et al. submitted). In addition, Pulau Karimata contains a fifth convergently evolved dwarf population, which is more closely related to a SW Bornean colugo than to any other

colugo population. Thus, we hypothesize that the causative genetic variants for dwarfism are likely different between Bornean colugos and Peninsular Malaysian / Sumatran colugos due to the millions of years of divergence between these populations. The recent independent origins and rapid change in morphology suggests that body size variation is a highly plastic phenotype that can rapidly change in response to new environmental conditions (Nagel & Schluter 1998).

The genetic similarities between colugos on major landmasses and their satellite islands revealed a potentially complex former dispersal corridor between Peninsular Malaysia and West Borneo across the Anambas and Natuna island chains. The Anambas islands lie east of Peninsular Malaysia and are currently isolated by a large span of shallow sea. The greater (Pulau Bunguran) and lesser (Pulau Serasan and Pulau Subi-besar) Natuna islands lie NW of Borneo. The Anambas islands are genetically similar to Peninsular Malaysia, the lesser Natuna islands are similar to W. Borneo. Somewhat surprisingly, the single colugo sample from the greater Natuna island, Pulau Bunguran appears to have recently diverged from the NE Bornean colugo population. Similarly, there is very little mtDNA divergence between greater mouse deer from Pulau Bunguran and Pulau Serasan and Peninsular Malaysian + Sumatran (1.4%) and W. Bornean (1.9%) greater mouse deer. Additionally, a Sunda pangolin from Pulau Bunguran is only 0.4% and 0.6% diverged from Peninsular Malaysian + Sumatran and W. Bornean sundaic pangolins respectively. These findings suggests that these islands acted as a crossroad for SE Asian mammalian taxa that moved between Borneo and the Indochinese Peninsula. These islands may have acted as biotic refuges for many mammals during

rising (isolating) sea levels, as has been proposed in frogs and snakes (Inger & Voris 2001). A small reduction in sea level (30-55 m) was sufficient to connect the lesser Natunas to NW Borneo and the Anambas islands to Peninsular Malaysia, respectively. However, access to the greater Natuna islands would have required a larger sea level drop (-75m below present levels) (Voris 2000, Sathiamurthy & Voris 2006). This indicates that dispersal and gene flow to Bunguran was probably less frequent than to the Anambas and lesser Natuna islands, and may explain the presence of endemic species on Bunguran such as the Natuna leaf monkey (Lammertink et al. 2003).

4.5 Conclusions and Future Research

The limited dispersal, structured phylogenies, and maternal and biparental evolutionary histories of colugos allowed us to decipher complex biogeographic patterns within southeast Asian mammals and generate hypotheses concerning the past distribution of mammalian taxa and, by extension, the ecological history and forested distributions of Sundaland. The low genetic divergence and close genetic relationship between colugos on adjacent landmasses provides substantial evidence that past forest distributions recently connected currently isolated islands and were far more expansive than at present. Although colugos were capable of moving to nearby landmasses recently through forested corridors we find that colugo populations separated by large geographic distances, such as colugos on Peninsular Malaysia and W. Borneo, diverged from one another in the Late Pliocene, but show little evidence for large distance dispersal in the Pleistocene despite the many glacial periods connecting these landmasses. Therefore we propose that forests present in the middle Pliocene (~4.5-3.5 Mya) might have been

more evergreen (less seasonal) with higher density canopies (>95%) and subsequently more suitable habitat for colugos, while forests present in the Pleistocene were more consistently fragmented, transient, seasonal, or had lower density canopies that prevented genetic exchange between colugo populations. Our results further suggest that significant cryptic species diversity is present among Sundaic colugos and lesser mouse deer populations. By contrast, little genetic diversity and phylogenetic structure was found within the greater mouse deer from a large geographic sampling from Peninsular Malaysia to E. Borneo, suggesting that they are more capable of long-distance dispersal than lesser mouse deer. Pangolins showed minimal phylogeographic structuring across these same regions, with the exception of E. Borneo and Western Sundaic regions, potentially indicating that montane regions in Borneo present more formidable barriers to dispersal than expanses of savanna that previously separated Borneo from Peninsular Malaysia / Sumatra during glacial periods. Nonetheless, broader sampling of populations from these species will improve the geographic discrimination of putative novel species boundaries identified here. Finally, sampling of other low vagility mammals may confirm and refine the broader phylogeographic patterns identified within Borneo, and reveal specific ecological barriers that hinder gene flow across this large and ecologically diverse island.

CHAPTER V

CONCLUSIONS

The objectives of this dissertation were to: 1) to develop methods to enrich for orthologous sequences from degraded, damaged, and contaminated sources of DNA that can be used for analysis with next-generation sequencing technologies (Chapter II), 2) to describe the genetic variation within order Dermoptera through comparisons of maternal, paternal, and biparental evolutionary histories, and determine appropriate species classifications for colugos (Chapter III), and 3) to describe how the current standing genetic variation in colugos became distributed across the Southeast Asian mainland and associated archipelagos (Chapter IV). The scope of this dissertation was broadend greatly subsequent to the sequencing of the colugo genome (Mason et al. submitted). This genome sequence enabled us to capture nuclear DNA sequences, perform the first genome-wide phylogenetic comparisons to definitively place the order Dermoptera as the sister lineage to Primates, and calculate the first molecular estimate of divergence time between the *Galeopterus* and *Cynocephalus* by comparing genome wide protein coding orthologs.

5.1 DNA Capture Hybridization

Enrichment for mtDNA genome sequences from museum specimens proved to be very efficient when coupled with mitochondrial organelle enrichment prior to DNA extraction and our touchdown hybridization procedure (Chapter II). Captured sequences were unevenly distributed across the mitochondrial genome suggesting that bias in probe

selection of target (museum) DNA occurs during the hybridization reaction. Our hypothesis that efficiency of DNA hybridization would decrease inversely with DNA sequence divergence between probe and target DNA sources was not substantiated by strong correlations. However, we did not remove PCR duplicates prior to this analysis which is certainly a counding factor. The exact cause of bias in DNA hybridization remains unknown.

Nuclear capture (Chapter III) proved to be much less inefficient than mtDNA capture. This was expected given the much higher abundance of mitochondrion (hundreds to thousands) relative to the nuclear genome sequences (two copies) per cell. Nonetheless, we successfully captured low coverage orthologous sequences for 81 kb of autosomal, 34 kb of neutral X-chromosome, and 24kb of Y-chromosome sequences that were suitable for phylogenetic comparisons.

5.2 Divergence Dating

Point estimates for divergence times between phylogenetic lineages are referenced throughout Chapters III, IV, and IV. Intraspecific divergence times for colugos, mouse deer, and pangolins are based upon one terminal molecular calibration. We acknowledge that our point estimates between colugo, mouse deer, and pangolins are potentially inaccurate to some degree due to the use secondary molecular callirations without fossil callibrations, however these point estimates are likely still informative as error introduced through use of secondary callibrations seems predictable (Schenk 2016). We propose that our point estimates represent rough approximations for times of divergence between phylogenetic lineages. and stronger emphasis should be given to the

divergence time confidence intervals rather than the point estimates. However, we think our methods of maximizing uncertainty in confidence intervals (Dos Reis 2014) are appropriate given the limited fossil calibrations.

5.3 Phylogeography

Phylogenies of mtDNA, biparental, and Y-chromosome sequences establish the first in depth look into colugo evolutionary histories across their range. Dispersal of colugos is limited to specific environments by their anatomical adaptations, which is consistent with the strong phylogenetic sorting of colugo individuals by geography, where colugos are most similar to those on the same island and phylogenetically sister to those of geographically close neighboring islands. A notable exception is the islands of Sumatra and Java. The distance between them is similar to that of Sumatra and Peninsular Malaysia, however Sumatran colugos are more closely related to those of Peninsular Malaysia (~0.52 Mya) than Java (~3.9 Mya).This could be the result of Sumatra and Java being connected for shorter periods of time than Sumatra and Peninsular Malaysia. However, this seems to be a poor explanation as Java and Peninsular Malaysia are completely isolated from Sumatra at comparable sea levels (-35m and -25m respectively), whereas connections between Sumatra and Java become fragmented earlier starting with sea levels at -50m (Sathiamurhy & Voris 2006).

Ecological barriers such as a savanna were probably present between Sundaic islands for various amounts of time (Bird 2005, Harrison 2006). Our data agree that if savanna corridors where present they were present with different intensities at different times in different regions. Colugos on Peninsular Malaysia and the Anambas islands

were completely isolated by shallow seas earlier (-50m) and across a larger geographic distance than colugos on Sumatra and Java, but diverged much more recently (<0.52 Mya) than colugos on Sumatra and Java (~3.9 Mya). Based on their restriction to forested environments, we we can hypothesize that the environment east of Sumatra and north of Java was a more substantial of a barrier to colugos more recently than the savanna (or lack thereof) between Peninsular Malaysia and the Anambas islands. Further evidence for this is the low genetic divergence for the greater mouse deer and sunda pangolin (<1.9% and <0.6%, respectively) across the Anambas and Natuna island chains; even though these species appear much less restricted than colugos by environmental conditions.

The ancestral area reconstructions (Chapter IV) predict that Sumatra and Java were colonized from a 'rapidly' dispersing ancestral mainland colugo population. However, Java was also colonized by colugos from an ancestral E. Bornean population, resulting in mitochondrial capture of the E. Bornean-like mitochondrial genome against an ancestral mainland nuclear genome. Hybridization through secondary contact of Javan colugos with Sumatran (or other Sundaic) colugos might have been prevented due to cyto-nuclear incompatabilities which have been shown to play a role in speciation (Burton et al. 2013). We hypothesize that cyto-nuclear incompatabilities could have been the cause of speciation between Sumatran and Javan colugos after capture of the E. Bornean like mtDNA genome. The exceedingly high mtDNA genetic divergence between modern Javan and Sumatran colugos (~9 Mya) suggest that cytonuclear

interactions would breakdown resulting in speciation, however ecological barriers could have also limited gene flow between these islands.

5.4 Future Directions

Primarily this dissertation was designed to answer (and generate) questions specific to colugo speciation and biogeography. However, the mtDNA data we generated from Sundaic mouse deer and pangolins enabled us to compare and contrast different species specific traits to hypothesize how different dispersal capabilities and ecology could limit or promote gene flow (and introgression) between populations in different regions. We identified significant genetic differentiation across Borneo that suggest colugos within this large island are differentiated into three or more species based upon concordant mtDNA and nuclear DNA signals. However, improved sampling is necessary to better delineate colugo species boundaries on this large island. Furthermore, mitochondrial evidence from populations in southwestern and southeastern Borneo show marked sequence divergence from other Bornean populations (Chapter III), but these currently lack nuclear data to substantiate this large mtDNA differentiation. In addition, no genetic data has been retrieved for colugos from northcentral Borneo or southcentral Borneo, and the Bornean interior has remained completely unsampled. Expanded sampling of colugos, and sampling of isolated mountainous regions within Borneo could reveal further diversity in this poorly documented species. We believe that broader samplings of low vagility mammals within Borneo would also be valuable to identify shared patterns of phylogenetic differentiation across eastern and western Borneo, as has

been observed in avian taxa (Moyle 2005, Sheldon 2009, Sheldon 2015) and now in colugos, lesser mouse deer, and pangolins.

Sampling of multiple individuals colugos, greater mouse deer, and sundaic pangolins from the Rhio archipelago, the Anambas and Natuna island chains, and larger landmasses of Peninsular Malaysia, Sumatra, and W. Borneo could unveil a complex and dynamic corridor for genetic exchange between Peninsular Malaysia/Sumatra and W. Borneo. Island populations serve as genetic 'vaults' that have stored genetic variants from past dispersal events making them true relics that could help address outstanding biogeographic questions about the timing and presence and / or absence of a savannah corridor through this region. The current distribution of colugos and their genetic similarity to neighboring landmasses indicates that past forest distributions were recently much larger than the present largely refugial and rapidly shrinking forest distributions.

5.5 Recommended Taxonomic Revisions

Perhaps the most noteworthy conclusion from this dissertation is that the current taxonomy of *Galeopterus* is a very poor representation of the genetic variation across extant colugo populations. We propose that a minimum of six species be classified within *Galeopterus*, and propose that at least two of the four currently recognized subspecies (Table 1.1) within *Galeopterus* be elevated to species level (Table 5.1) 1) *G. variegatus* to represent colugos on Java and 2) *G. temmincki* to represent Peninsular Malaysian, Thailand, and Sumatran colugos. Subspecies *G. v. peninsulae* should be changed to *G. temmincki peninsulae*, and subspecies *G. v. temmincki* should be changed to *G. temmincki temmincki*. *G. borneanus* should be proposed for E. Bornean colugos

(Table 1.1, Table 5.1). However the subspecies (*G. v. borneanus*) was described based on colugos from SE. Borneo (Lyon 1911) and the type specimen from for E. Bornean colugos might better represent this clade as *G. lechei*, because *G. v. lechei* describes colugos from Central E. Borneo, however this nomenclature was proposed after *G . v. borneanus* by Gyldenstolpe in 1920. Colugos from NE. Borneo likely requires new formal taxonomic classification, however the most appropriate previously proposed nomenclature would likely be *G. natunae* (Miller 1903) or *G. hantu* (Cabrera 1924). Additionally, if nuclear evidence supports differentiation of SE. Bornean colugos from E. Bornean colugos *G. v. lautensis* could be elevated to species level classification (Lyon1911) (Table 5.1). However, we propose that subspecies *G. v. lautensis* could be an accepted subspecies based upon mtDNA genetic divergence of GVA_49 from E. Bornean colugos. Colugos from West Borneo should be represented by *G. gracilis* (proposed by Miller in 1903) and likely not *G. abotti* (proposed by Lyon in 1911). Colugos from Vietnam and Laos require new formal taxonomic classification.

Table 5.1. Proposed taxonomic revisions for *Galeopterus*. Nomenclature: A = currently accepted, P = previously proposed.

#	Species	A/P	Subspecies	Geographic Location
1	*G. variegatus*	A	*G. v. variegatus*	Java
2	*G. temmincki*	A	*G. t. peninsulae*	Malay States
3	*G. temmincki*	A	*G. t. temmincki*	Sumatra
4	*G. borneanus*	A	*G. b. borneanus*	E. Borneo
5	*G. gracilis*	P	*G. g. abotti*	Penebangan Indonesia, W. Borneo
6	*G. gracilis*	P	*G. g. gracilis*	Pulau Serasan (or Sirhassen)
7	*G. borneanus*	P	*G. b. lechei*	Central E. Borneo

Table 5.1 Continued.

#	Species	A/P	Subspecies	Geographic Location
8	New nomenclature or *G. natunae* or *G. hantu*	P	New nomenclature or *G. n. hantu* or *G. h. hantu*	North Sarawak, Borneo
9	*G. borneanus*	A	*G. b. lautensis*	Pulo Laut Indonesia, Pulau Sebuku, SE. Borneo
10	New nomenclature or *G. natunae* or *G. hantu*	P	New nomenclature or *G. n. natunae* or *G. h. natunae*	Pulau Bunguran
11	*G. temmincki*	P	*G. v. perhentianus*	East Perhentian Island
12	*G. temmincki*	P	*G. v. chombolis*	Pulau Chombol, Rhio Archipelago
13	*G. temmincki*	P	*G. v. taylori*	Pulau Tiomon
14	*G. temmincki*	P	*G. v. aoris*	Pulau Aur (or Aor)
15	*G. temmincki*	P	*G. v. terutaus*	Pulau Terutau
16	*G. temmincki*	P	*G. v. pumilus*	Pulau Adang
17	*G. variegatus*? or *G. temmincki?*	P	*G. v. undatus*	Sumatra? / Java?
18	*G. temmincki*	P	*G. v. saturatus*	Pulau Tanah bala, Batu Islands
19	*G. temmincki*	P	*G. v. tellonis*	Pulau Tello, Batu Islands
20	*G. temmincki*	P	*G. v. tuancus*	Pulau Tuangku, Banjak Islands
21	New nomenclature Laos and Vietnam	A	New nomenclature Laos	Laos
22	New nomenclature Laos and Vietnam	A	New Nomenclature Vietnam	Vietnam

REFERENCES

Achmadi, A.S., Esselstyn, J.A., Rowe, K.C., Maryanto, I. & Abdullah, M. (2013) Phylogeny, diversity, and biogeography of Southeast Asian spiny rats (Maxomys). *Journal of Mammalogy* **94,** 1412-1423.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215,** 403-410.

Ammerman, L.K. & Hillis, D.M. (1992) A molecular test of bat relationships: monophyly or diphyly? *Systematic Biology* **41,** 222-232.

Arnason, U., Adegoke, J.A., Bodin, K., Born, E.W., Esa, Y.B., Gullberg, A., Nilsson, M., Short, R.V., Xu, X. & Janke, A. (2002) Mammalian mitogenomic relationships and the root of the eutherian tree. *Proceedings of the National Academy of Sciences* **99,** 8151-8156.

Audebert, J. (1799) Histoire naturelle des singes ei des makis.

Baker, R.J. & Bradley, R.D. (2006) Speciation in mammals and the genetic species concept. *Journal of Mammalogy* **87,** 643-662.

Barton, R.A., Purvis, A. & Harvey, P.H. (1995) Evolutionary radiation of visual and olfactory brain systems in primates, bats and insectivores. *Philosophical transactions of the Royal Society of London.Series B, Biological Sciences* **348,** 381-392.

Beard, K.C. (1993) Origin and evolution of gliding in early Cenozoic Dermoptera (Mammalia, Primatomorpha). *Primates and their relatives in phylogenetic perspective,* pp. 63-90. Springer.

Beard, K. (1991) Vertical postures and climbing in the morphotype of Primatomorpha: Implications for locomotor evolution in primate history. *Origines de la Bipédie chez les Hominidés* 79-87.

Benton, M.J., Donoghue, P.C., Asher, R.J., Friedman, M., Near, T.J. & Vinther, J. (2015) Constraints on the timescale of animal evolutionary history. *Palaeontologia Electronica* **18,** 1-106.

Benton, M., Donoghue, P. & Asher, R. (2009) Calibrating and constraining molecular clocks. *The Timetree of Life* 35-86.

Binladen, J., Wiuf, C., Gilbert, M.T.P., Bunce, M., Barnett, R., Larson, G., Greenwood, A.D., Haile, J., Ho, S.Y. & Hansen, A.J. (2006) Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. *Genetics* **172,** 733-741.

Bird, M.I., Taylor, D. & Hunt, C. (2005) Palaeoenvironments of insular Southeast Asia during the Last Glacial Period: a savanna corridor in Sundaland? *Quaternary Science Reviews* **24,** 2228-2242.

Bloch, J.I., Silcox, M.T., Boyer, D.M. & Sargis, E.J. (2007) New Paleocene skeletons and the relationship of plesiadapiforms to crown-clade primates. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 1159-1164.

Boddaërt, P. (1768) Dierkundig mengelwerk in het welke de nieuwe of nog duistere zoorten van dieren. *Abraham van Paddenburg en J. Van Schoonhoven* **2**.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D. & Pirovano, W. (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics (Oxford, England)* **27,** 578-579.

Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Schmitz, R., Doronichev, V.B., Golovanova, L.V., de la Rasilla, M., Fortea, J., Rosas, A. & Paabo, S. (2009) Targeted retrieval and analysis of five neandertal mtDNA genomes. *Science (New York, N.Y.)* **325,** 318-321.

Brown, R.M., Weghorst, J.A., Olson, K.V., Duya, M.R., Barley, A.J., Duya, M.V., Shekelle, M., Neri-Arboleda, I., Esselstyn, J.A. & Dominy, N.J. (2014) Conservation genetics of the Philippine Tarsier: cryptic genetic variation restructures conservation priorities for an island archipelago primate. *PloS One* **9,** e104340.

Brown, R. & Diesmos, A. (2009) Philippines, biology. *Encyclopedia of Islands.University of California Press, Berkeley* 723-732.

Burton, R.S., Pereira, R.J. & Barreto, F.S. (2013) Cytonuclear genomic interactions and hybrid breakdown. *Annual Review of Ecology, Evolution, and Systematics* **44,** 281-302.

Byrnes, G., Lim, N.T. & Spence, A.J. (2008) Take-off and landing kinetics of a free-ranging gliding mammal, the Malayan colugo (Galeopterus variegatus). *Proceedings.Biological sciences / The Royal Society* **275,** 1007-1013.

Cabrera, A. & Museo nacional de ciencias naturales (Madrid). (1925) *Genera Mammalium: Insectivora Galeopithecia*. Museo nacional de ciencias naturales.

Cannon, C.H., Morley, R.J. & Bush, A.B. (2009) The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 11188-11193.

Chaisson, M.J. & Pevzner, P.A. (2008) Short read fragment assembly of bacterial genomes. *Genome Research* **18,** 324-330.

Chasen, F. & Kloss, C.B. (1929) Notes on flying lemurs (Galeopterus). *Bulletin of the Raffles Museum* **2,** 12-22.

Chen, K., Durand, D. & Farach-Colton, M. (2000) NOTUNG: a program for dating gene duplications and optimizing gene family trees. *Journal of Computational Biology* **7,** 429-447.

Chester, S.G., Bloch, J.I., Boyer, D.M. & Clemens, W.A. (2015) Oldest known euarchontan tarsals and affinities of Paleocene Purgatorius to Primates. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 1487-1492.

Coyne, J.A. & Orr, H.A. (2004) *Speciation.* Sinauer Associates Sunderland, MA.

Cracraft, J. (1997) Species concepts in systematics and conservation biology - an ornithological viewpoint. *Species: The Units of Biodiversity* (ed. by M. F. Claridge, H. A. Dawah, and M. R. Wilson), pp. 325-325-340. Chapman & Hall, London, UK.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.L. & Meyer, M. (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 15758-15763.

de Bruyn, M., Stelbrink, B., Morley, R.J., Hall, R., Carvalho, G.R., Cannon, C.H., van den Bergh, G., Meijaard, E., Metcalfe, I., Boitani, L., Maiorano, L., Shoup, R. & von Rintelen, T. (2014) Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity. *Systematic Biology* **63,** 879-901.

Del Mastro, R.G. & Lovett, M. (1997) Isolation of coding sequences from genomic regions using direct selection. *Gene Isolation and Mapping Protocols*, pp. 183-199. Springer.

den Tex, R., Thorington, R., Maldonado, J.E. & Leonard, J.A. (2010) Speciation dynamics in the SE Asian tropics: putting a time perspective on the phylogeny and biogeography of Sundaland tree squirrels, Sundasciurus. *Molecular Phylogenetics and Evolution* **55,** 711-720.

Dos Reis, M., Zhu, T. & Yang, Z. (2014) The impact of the rate prior on Bayesian estimation of divergence times with multiple Loci. *Systematic Biology* **63,** 555-565.

Douzery, E.J., Scornavacca, C., Romiguier, J., Belkhir, K., Galtier, N., Delsuc, F. & Ranwez, V. (2014) OrthoMaM v8: a database of orthologous exons and coding sequences for comparative genomics in mammals. *Molecular Biology and Evolution* **31,** 1923-1928.

Dunham, A.E., Tinkle, D.W. & Gibbons, J.W. (1978) Body size in island lizards: a cautionary tale. *Ecology* 1230-1238.

Emerling, C.A., Huynh, H.T., Nguyen, M.A., Meredith, R.W. & Springer, M.S. (2015) Spectral shifts of mammalian ultraviolet-sensitive pigments (short wavelength-sensitive opsin 1) are associated with eye length and photic niche evolution. *Proceedings.Biological sciences / The Royal Society* **282,** 10.1098/rspb.2015.1817.

Endo, H., Fukuta, K., Kimura, J., Sasaki, M., Hayashi, Y. & Oshida, T. (2004) Phylogenetic relationships among populations of the mouse deer in the Southeast Asian Region from the nucleotide sequence of cytochrome b gene. *Mammal Study* **29,** 119-123.

Excoffier, L. & Lischer, H.E. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10,** 564-567.

Fan, Y., Huang, Z., Cao, C., Chen, C., Chen, Y., Fan, D., He, J., Hou, H., Hu, L. & Hu, X. (2013) Genome of the Chinese tree shrew. *Nature Communications* **4,** 1426.

Gilbert, M.T., Tomsho, L.P., Rendulic, S., Packard, M., Drautz, D.I., Sher, A., Tikhonov, A., Dalen, L., Kuznetsova, T., Kosintsev, P., Campos, P.F., Higham, T., Collins, M.J., Wilson, A.S., Shidlovskiy, F., Buigues, B., Ericson, P.G., Germonpre, M., Gotherstrom, A., Iacumin, P., Nikolaev, V., Nowak-Kemp, M., Willerslev, E., Knight, J.R., Irzyk, G.P., Perbost, C.S., Fredrikson, K.M., Harkins, T.T., Sheridan, S., Miller, W. & Schuster, S.C. (2007) Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science (New York, N.Y.)* **317,** 1927-1930.

Gorog, A.J., Sinaga, M.H. & Engstrom, M.D. (2004) Vicariance or dispersal? Historical biogeography of three Sunda shelf murine rodents (Maxomys surifer, Leopoldamys sabanus and Maxomys whiteheadi). *Biological Journal of the Linnean Society* **81,** 91-109.

Gregory, W.K. (1910) *The orders of mammals.* The Trustees.

Gunnell, G.F. & Simmons, N.B. (2005) Fossil evidence and the origin of bats. *Journal of Mammalian Evolution* **12,** 209-246.

Haldane, J.B. (1922) Sex ratio and unisexual sterility in hybrid animals. *Journal of genetics* **12,** 101-109.

Hall, R. (2013) The palaeogeography of Sundaland and Wallacea since the Late Jurassic. *Journal of Limnology* **72,** 1.

Hall, R. (2002) Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. *Journal of Asian Earth Sciences* **20,** 353-431.

Hall, R., Ali, J.R., Anderson, C.D. & Baker, S.J. (1995) Origin and motion history of the Philippine Sea Plate. *Tectonophysics* **251,** 229-250.

Han, K. (2000) Phylogeny and biogeography of tree shrews (Scandentia: Tupaiidae). *Biological Sciences.Louisiana State University, Baton Rouge* .

Haq, B.U., Hardenbol, J. & Vail, P.R. (1987) Chronology of fluctuating sea levels since the triassic. *Science (New York, N.Y.)* **235,** 1156-1167.

Harrison, T., Krigbaum, J. & Manser, J. (2006) Primate biogeography and ecology on the Sunda Shelf islands: a paleontological and zooarchaeological perspective. *Primate Biogeography* (ed. by Anonymous), pp. 331-372. Springer.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., van Vuuren, B.J., Matthee, C., Ruiz-Garcia, M., Catzeflis, F., Areskoug, V. & Nguyen, T.T. (2012) Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes Rendus Biologies* **335,** 32-50.

Hawkins, M.T., Hofman, C.A., Callicrate, T., McDonough, M.M., Tsuchiya, M.T., Gutiérrez, E.E., Helgen, K.M. & Maldonado, J.E. (2015) In-solution hybridization for mammalian mitogenome enrichment: pros, cons and challenges associated with multiplexing degraded DNA. *Molecular Ecology Resources*.

Hawkins, R.D., Hon, G.C. & Ren, B. (2010) Next-generation genomics: an integrative approach. *Nature Reviews Genetics* **11,** 476-486.

Hazkani-Covo, E., Zeller, R.M. & Martin, W. (2010) Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet* **6,** e1000834.

Heaney, L.R. (1991) A synopsis of climatic and vegetational change in Southeast Asia. *Climatic Change* **19,** 53-61.

Heaney, L.R. (1978) Island area and body size of insular mammals: evidence from the tri-colored squirrel (Callosciurus prevosti) of Southeast Asia. *Evolution* 29-44.

Hedges, S.B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. (2015) Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution* **32,** 835-845.

Hope, G.S. (2004) Glaciation of Malaysia and Indonesia, excluding New Guinea. *Quaternary Glaciations—Extent and Chronology, Part III: South America, Asia, Africa, Australasia, Antarctica.Developments in Quaternary Science, 2c.Elsevier, Amsterdam* 211-214.

Hope, G. (2007) Paleoecology and paleoenvironments of Papua. *Andrew J. Marshall & Bruce M. Beehler (ed.), The Ecology of Papua: Part One* **Periplus,** 255-266.

Huang, X. & Madan, A. (1999) CAP3: A DNA sequence assembly program. *Genome research* **9,** 868-877.

Inger, R.F. & Voris, H.K. (2001) The biogeographical relations of the frogs and snakes of Sundaland. *Journal of Biogeography* **28,** 863-891.

Jackson, S.M. & Thorington Jr, R.W. (2012) Gliding Mammals: Taxonomy of living and extinct species. *Smithsonian Contributions to Zoology* .

Janečka, J.E., Helgen, K.M., Lim, N.T., Baba, M., Izawa, M. & Murphy, W.J. (2008) Evidence for multiple species of Sunda colugo. *Current Biology* **18,** R1001-R1002.

Janecka, J.E., Miller, W., Pringle, T.H., Wiens, F., Zitzmann, A., Helgen, K.M., Springer, M.S. & Murphy, W.J. (2007) Molecular and genomic data identify the closest living relative of Primates. *Science* **318,** 792-794.

Jones, C., Tegelström, H., Latchman, D. & Berry, R. (1988) An improved rapid method for mitochondrial DNA isolation suitable for use in the study of closely related populations. *Biochemical Genetics* **26,** 83-88.

Kapur, V.V. & Bajpai, S. Oldest South Asian tapiromorph (Perissodactyla, Mammalia) from the Cambay Shale Formation, western India, with comments on its phylogenetic position and biogeographic implications.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30,** 772-780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)* **28,** 1647-1649.

Kominz, M., Browning, J., Miller, K., Sugarman, P., Mizintseva, S. & Scotese, C. (2008) Late Cretaceous to Miocene sea-level estimates from the New Jersey and Delaware coastal plain coreholes: An error analysis. *Basin Research* **20,** 211-226.

Kriegs, J.O., Churakov, G., Jurka, J., Brosius, J. & Schmitz, J. (2007) Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends in Genetics* **23,** 158-161.

Kriegs, J.O., Churakov, G., Kiefmann, M., Jordan, U., Brosius, J. & Schmitz, J. (2006) Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biology* **4,** e91.

Kumar, S., Stecher, G., Peterson, D. & Tamura, K. (2012) MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics (Oxford, England)* **28,** 2685-2686.

Kuritzin, A., Kischka, T., Schmitz, J., Churakov, G. (in press) Incomplete Lineage Sorting and hybridization statistics for retroposon insertion data. *PLOS Comp. Biology.*

Lammertink, M., Nijman, V. & Setiorini, U. (2003) Population size, Red List status and conservation of the Natuna leaf monkey Presbytis natunae endemic to the island of Bunguran, Indonesia. *Oryx* **37,** 472-479.

Larson, G., Albarella, U., Dobney, K., Rowley-Conwy, P., Schibler, J., Tresset, A., Vigne, J.D., Edwards, C.J., Schlumbaum, A., Dinu, A., Balacsescu, A., Dolman,

G., Tagliacozzo, A., Manaseryan, N., Miracle, P., Van Wijngaarden-Bakker, L., Masseti, M., Bradley, D.G. & Cooper, A. (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 15276-15281.

Leche, W. (1886) *Über die säugethiergattung Galeopithecus: eine morphologische untersuchung.* PA Norstedt & söner.

Leonard, J.A., Tex, R., Hawkins, M.T., Muñoz-Fuentes, V., Thorington, R. & Maldonado, J.E. (2015) Phylogeography of vertebrates on the Sunda Shelf: a multi-species comparison. *Journal of Biogeography* **42,** 871-879.

Li, G., Davis, B.W., Eizirik, E., Murphy, W.J., Lin, C., Taggart, A.J., Lim, K.H., Cygan, K.J., Ferraris, L. & Creton, R. (2016) Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Research* **26,** 1-11.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25,** 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & 1000 Genome Project Data Processing Subgroup. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England)* **25,** 2078-2079.

Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)* **25,** 1451-1452.

Lim, N.T. & Ng, P.K. (2007) Home range, activity cycle and natal den usage of a female Sunda pangolin Manis javanica (Mammalia: Pholidota) in Singapore. *Endangered Species Research* **3,** 1-8.

Lim, N.T., Giam, X., Byrnes, G. & Clements, G.R. (2013) Occurrence of the Sunda colugo (Galeopterus variegatus) in the tropical forests of Singapore: A Bayesian approach. *Mammalian Biology-Zeitschrift für Säugetierkunde* **78,** 63-67.

Lin, J., Chen, G., Gu, L., Shen, Y., Zheng, M., Zheng, W., Hu, X., Zhang, X., Qiu, Y. & Liu, X. (2014) Phylogenetic affinity of tree shrews to Glires is attributed to fast evolution rate. *Molecular Phylogenetics and Evolution* **71,** 193-200.

Linnaeus, C. (1758) Systema naturae, vol. 1. *Systema Naturae, Vol.1*.

Lohman, D.J., de Bruyn, M., Page, T., von Rintelen, K., Hall, R., Ng, P.K., Shih, H., Carvalho, G.R. & von Rintelen, T. (2011) Biogeography of the Indo-Australian archipelago. *Annual Review of Ecology, Evolution, and Systematics* **42,** 205-226.

Löytynoja, A. (2014) Phylogeny-aware alignment with PRANK. *Multiple Sequence Alignment Methods* 155-170.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.W. & Wang, J. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1,** 18-217X-1-18.

Lyon, M.W. (1908) XXI.—On a collection of mammals from the Batu Islands, west of Sumatra. *Journal of Natural History* **1,** 137-140.

Lyon, M.W. (1911) Mammals collected by Dr. WL Abbott on Borneo and some of the small adjacent islands. *Proceedings of the US National Museum* **40,** 53-156.

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J. & Turner, D.J. (2010) Target-enrichment strategies for next-generation sequencing. *Nature Methods* **7,** 111-118.

Maricic, T., Whitten, M. & Pääbo, S. (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PloS One* **5,** e14004.

Maseko, B.C., Bourne, J.A. & Manger, P.R. (2007) Distribution and morphology of cholinergic, putative catecholaminergic and serotonergic neurons in the brain of the Egyptian Rousette flying fox, Rousettus aegyptiacus. *Journal of Chemical Neuroanatomy* **34,** 108-127.

Mason, V.C., Li, G., Helgen, K.M. & Murphy, W.J. (2011) Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Research* **21,** 1695-1704.

Matzke, N.J. (2013) BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. R package, version 0.2.1.

Matzke, N.J. (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in Island Clades. *Systematic Biology* **63,** 951-970.

Mayr, E. (1942) *Systematics and the origin of species, from the viewpoint of a zoologist.* Harvard University Press.

Meijaard, E. (2004) Solving mammalian riddles: a reconstruction of the Tertiary and Quaternary distribution of mammals and their palaeoenvironments in island South-East Asia.

Meijaard, E. (2003) Mammals of south-east Asian islands and their Late Pleistocene environments. *Journal of Biogeography* **30,** 1245-1257.

Meijaard, E. & Groves, C. (2004) A taxonomic revision of the Tragulus mouse-deer (Artiodactyla). *Zoological Journal of the Linnean Society* **140,** 63-102.

Melin, A.D., Wells, K., Moritz, G.L., Kistler, L., Orkin, J.D., Timm, R.M., Bernard, H., Lakim, M.B., Perry, G.H., Kawamura, S. & Dominy, N.J. (2016) Euarchontan opsin variation brings new focus to Primate origins. *Molecular Biology and Evolution*.

Meredith, R.W., Janečka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A., Eizirik, E., Simão, T.L. & Stadler, T. (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334,** 521-524.

Metcalfe, I. (2011) Tectonic framework and Phanerozoic evolution of Sundaland. *Gondwana Research* **19,** 3-21.

Millar, C.D., Huynen, L., Subramanian, S., Mohandesan, E. & Lambert, D.M. (2008) New developments in ancient genomics. *Trends in Ecology & Evolution* **23,** 386-393.

Miller, G.S. (1903) *Seventy new Malayan mammals*, 45:1-55 edn. Smithsonian Institution.

Miller, G.S. (1900) *Mammals collected by Dr. WL Abbott on Pulo Lankawi and the Butang Islands*.

Miller, W., Drautz, D.I., Ratan, A., Pusey, B., Qi, J., Lesk, A.M., Tomsho, L.P., Packard, M.D., Zhao, F. & Sher, A. (2008) Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456,** 387-390.

Miller, W., Drautz, D.I., Janecka, J.E., Lesk, A.M., Ratan, A., Tomsho, L.P., Packard, M., Zhang, Y., McClellan, L.R., Qi, J., Zhao, F., Gilbert, M.T., Dalen, L., Arsuaga, J.L., Ericson, P.G., Huson, D.H., Helgen, K.M., Murphy, W.J.,

Gotherstrom, A. & Schuster, S.C. (2009) The mitochondrial genome sequence of the Tasmanian tiger (Thylacinus cynocephalus). *Genome Research* **19,** 213-220.

Montague, M.J., Li, G., Gandolfi, B., Khan, R., Aken, B.L., Searle, S.M., Minx, P., Hillier, L.W., Koboldt, D.C., Davis, B.W., Driscoll, C.A., Barr, C.S., Blackistone, K., Quilez, J., Lorente-Galdos, B., Marques-Bonet, T., Alkan, C., Thomas, G.W., Hahn, M.W., Menotti-Raymond, M., O'Brien, S.J., Wilson, R.K., Lyons, L.A., Murphy, W.J. & Warren, W.C. (2014) Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 17230-17235.

Moritz, C. (1994) Defining 'evolutionarily significant units' for conservation. *Trends in Ecology & Evolution* **9,** 373-375.

Moritz, G.L., Lim, N.T., Neitz, M., Peichl, L. & Dominy, N.J. (2013) Expression and evolution of short wavelength sensitive opsins in colugos: A nocturnal lineage that informs debate on primate origins. *Evolutionary Biology* 1-12.

Morley, R.J. (2000) *Origin and evolution of tropical rain forests.*John Wiley & Sons.

Moss, S.J. & Wilson, M.E. (1998) Biogeographic implications of the Tertiary palaeogeographic evolution of Sulawesi and Borneo. *Biogeography and Geological Evolution of SE Asia* 133-163.

Moyle, R.G., Schilthuizen, M., Rahman, M.A. & Sheldon, F.H. (2005) Molecular phylogenetic analysis of the white-crowned forktail Enicurus leschenaulti in Borneo. *Journal of Avian Biology* **36,** 96-101.

Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A. & O'Brien, S.J. (2001) Molecular phylogenetics and the origins of placental mammals. *Nature* **409,** 614-618.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature* **403,** 853-858.

Nagel, L. & Schluter, D. (1998) Body size, natural selection, and speciation in sticklebacks. *Evolution* 209-218.

Novacek, M.J. & Wyss, A.R. (1986) Higher-level relationships of the recent eutherian orders: morphological evidence. *Cladistics* **2,** 257-287.

O'Leary, M.A., Bloch, J.I., Flynn, J.J., Gaudin, T.J., Giallombardo, A., Giannini, N.P., Goldberg, S.L., Kraatz, B.P., Luo, Z.X., Meng, J., Ni, X., Novacek, M.J., Perini,

F.A., Randall, Z.S., Rougier, G.W., Sargis, E.J., Silcox, M.T., Simmons, N.B., Spaulding, M., Velazco, P.M., Weksler, M., Wible, J.R. & Cirranello, A.L. (2013) The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science (New York, N.Y.)* **339,** 662-667.

Patou, M., Wilting, A., Gaubert, P., Esselstyn, J.A., Cruaud, C., Jennings, A.P., Fickel, J. & Veron, G. (2010) Evolutionary history of the Paradoxurus palm civets–a new model for Asian biogeography. *Journal of Biogeography* **37,** 2077-2097.

Perelman, P., Johnson, W.E., Roos, C., Seuánez, H.N., Horvath, J.E., Moreira, M., Kessing, B., Pontius, J., Roelke, M. & Rumpler, Y. (2011) A molecular phylogeny of living primates. *PLoS Genetics* **7,** e1001342.

Peterson, A. (1998) Direct cDNA selection. *Genome analysis: A laboratory manual (ed.Birren B et al.)* 159-171.

Petit, R.J. & Excoffier, L. (2009) Gene flow and species delimitation. *Trends in Ecology & Evolution* **24,** 386-393.

Pettigrew, J.D. (1995) Flying primates: crashed, or crashed through? *Symposia of the Zoological Society London.* ***67,*** 3-26.

Pettigrew, J.D., Maseko, B. & Manger, P. (2008) Primate-like retinotectal decussation in an echolocating megabat, Rousettus aegyptiacus. *Neuroscience* **153,** 226-231.

Pettigrew, J.D., Jamieson, B., Robson, S., Hall, L., McAnally, K. & Cooper, H. (1989) Phylogenetic relations between microbats, megabats and primates (Mammalia: Chiroptera and Primates). *Philosophical Transactions of the Royal Society of London.Series B, Biological Sciences* 489-559.

Prothero, D.R. & Foss, S.E. (2007) *The evolution of artiodactyls.* JHU Press.

R Core Team. (2015) R: A language and environment for statistical computing. Vienna, Austria; 2014. *URL* http://www.R-project.org .

Raes, N., Cannon, C.H., Hijmans, R.J., Piessens, T., Saw, L.G., van Welzen, P.C. & Slik, J.W. (2014) Historical distribution of Sundaland's Dipterocarp rainforests at Quaternary glacial maxima. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 16790-16795.

Rafinesque, C.S. (1815) *Analyse de la nature, ou tableau de l'univers et des corps organisés*1815 aux dépens de l'auteur.

Ree, R. & Smith, S. (2008) Lagrange: software for likelihood analysis of geographic range evolution. *Systematic Biology* **57,** 4-14.

Ronquist, F. (1997) Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* **46,** 195-203.

Rose, K.D., DeLeon, V.B., Missiaen, P., Rana, R.S., Sahni, A., Singh, L. & Smith, T. (2008) Early Eocene lagomorph (Mammalia) from Western India and the early diversification of Lagomorpha. *Proceedings.Biological Sciences / The Royal Society* **275,** 1203-1208.

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

Schenk, J.J. (2016) Consequences of secondary calibrations on divergence time estimates. *PloS One* **11,** e0148228.

Schmitz, J., Ohme, M., Suryobroto, B. & Zischler, H. (2002) The Colugo (Cynocephalus variegatus, Dermoptera): The Primates' Gliding Sister? *Molecular Biology and Evolution* **19,** 2308-2312.

Sheldon, F.H., Lim, H.C. & Moyle, R.G. (2015) Return to the Malay Archipelago: the biogeography of Sundaic rainforest birds. *Journal of Ornithology* 1-23.

Sheldon, F.H., Lohman, D.J., Lim, H.C., Zou, F., Goodman, S.M., Prawiradilaga, D.M., Winker, K., Braile, T.M. & Moyle, R.G. (2009) Phylogeography of the magpie-robin species complex (Aves: Turdidae: Copsychus) reveals a Philippine species, an interesting isolating barrier and unusual dispersal patterns in the Indian Ocean and Southeast Asia. *Journal of Biogeography* **36,** 1070-1083.

Shoup[1], R.C., Morley, R.J., Swiecicki, T. & Clark, S. (2013) Tectono-stratigraphic framework and Tertiary paleogeogeography of Southeast Asia; Gulf of Thailand to South Vietnam Shelf. *Houston Geological Society Bulletin* **55**.

Simpson, G.G. (1945) The principles of classification and a classification of mammals. *Bull.Amer.Museum Nat.History.* **85,** .

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J. & Bieri, T. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423,** 825-837.

Smith, J.D. (1980) Chiropteran phylogenetics: introduction. 233-244.

Soh, Y.S., Alföldi, J., Pyntikova, T., Brown, L.G., Graves, T., Minx, P.J., Fulton, R.S., Kremitzki, C., Koutseva, N. & Mueller, J.L. (2014) Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* **159,** 800-813.

Stafford, B.J. (1999) *Taxonomy and ecological morphology of the flying lemurs (Dermoptera, Cynocephalidae).*

Stafford, B.J. & Szalay, F.S. (2000) Craniodental functional morphology and taxonomy of dermopterans. *Journal of Mammalogy* **81,** 360-385.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)* **30,** 1312-1313.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)* **22,** 2688-2690.

Steppan, S.J., Zawadzki, C. & Heaney, L.R. (2003) Molecular phylogeny of the endemic Philippine rodent Apomys (Muridae) and the dynamics of diversification in an oceanic archipelago. *Biological Journal of the Linnean Society* **80,** 699-715.

Storey, E. (2014) Genetic cerebellar ataxias. *Seminars in Neurology* **34,** 280-292.

Summerer, D. (2009) Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. *Genomics* **94,** 363-368.

Sussman, R.W., Tab Rasmussen, D. & Raven, P.H. (2013) Rethinking primate origins again. *American Journal of Primatology* **75,** 95-106.

Talavera, G. & Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56,** 564-577.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24,** 1596-1599.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30,** 2725-2729.

Teeling, E.C., Springer, M.S., Madsen, O., Bates, P., O'Brien, S.J. & Murphy, W.J. (2005) A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* **307,** 580-584.

Thomas, O. (1908) XXXIX.—The nomenclature of the flying-lemurs. *The Annals and Magazine of Natural History* **1,** 252-255.

Toews, D.P. & Brelsford, A. (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* **21,** 3907-3930.

Tosi, A.J. & Coke, C.S. (2007) Comparative phylogenetics offer new insights into the biogeographic history of Macaca fascicularis and the origin of the Mauritian macaques. *Molecular Phylogenetics and Evolution* **42,** 498-504.

Triant, D.A. & DeWoody, J.A. (2007) The occurrence, detection, and avoidance of mitochondrial DNA translocations in mammalian systematics and phylogeography. *Journal of Mammalogy* **88,** 908-920.

Uba, C.E., Strecker, M.R. & Schmitt, A.K. (2007) Increased sediment accumulation rates and climatic forcing in the central Andes during the late Miocene. *Geology* **35,** 979-982.

Waddell, P.J., Kishino, H. & Ota, R. (2001) A phylogenetic foundation for comparative mammalian genomics. *Genome Informatics* **12,** 141-154.

Wang, G., Shi, P., Zhu, Z. & Zhang, Y.P. (2010) More functional V1R genes occur in nest-living and nocturnal terricolous mammals. *Genome Biology and Evolution* **2,** 277-283.

Wilcove, D.S., Giam, X., Edwards, D.P., Fisher, B. & Koh, L.P. (2013) Navjot's nightmare revisited: logging, agriculture, and biodiversity in Southeast Asia. *Trends in Ecology & Evolution* **28,** 531-540.

Wilson, D.E. & Reeder, D.M. (2005) *Mammal Species of the World: a taxonomic and geographic reference*Johns Hopkins University Press.

Wischusen, E.W. & Richmond, M.E. (1998) Foraging ecology of the Philippine flying lemur (Cynocephalus volans). *Journal of Mammalogy* **79,** 1288-1295.

Woodruff, D.S. (2010) Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. *Biodiversity and Conservation* **19,** 919-941.

Yang, Z. (2015) The BPP program for species tree estimation and species delimitation. *Current Zoology* **61,** 854-865.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24,** 1586-1591.

You, F.M., Huo, N., Gu, Y.Q., Luo, M.C., Ma, Y., Hane, D., Lazo, G.R., Dvorak, J. & Anderson, O.D. (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC bioinformatics* **9,** 253-2105-9-253.

Zhang, H., Miller, M.P., Yang, F., Chan, H.K., Gaubert, P., Ades, G. & Fischer, G.A. (2015) Molecular tracing of confiscated pangolin scales for conservation and illegal trade monitoring in Southeast Asia. *Global Ecology and Conservation* **4,** 414-422.

Ziegler, T., Abegg, C., Meijaard, E., Perwitasari-Farajallah, D., Walter, L., Hodges, J.K. & Roos, C. (2007) Molecular phylogeny and evolutionary history of Southeast Asian macaques forming the M. silenus group. *Molecular Phylogenetics and Evolution* **42,** 807-816.

APPENDIX A

SUPPLEMENTAL FIGURES

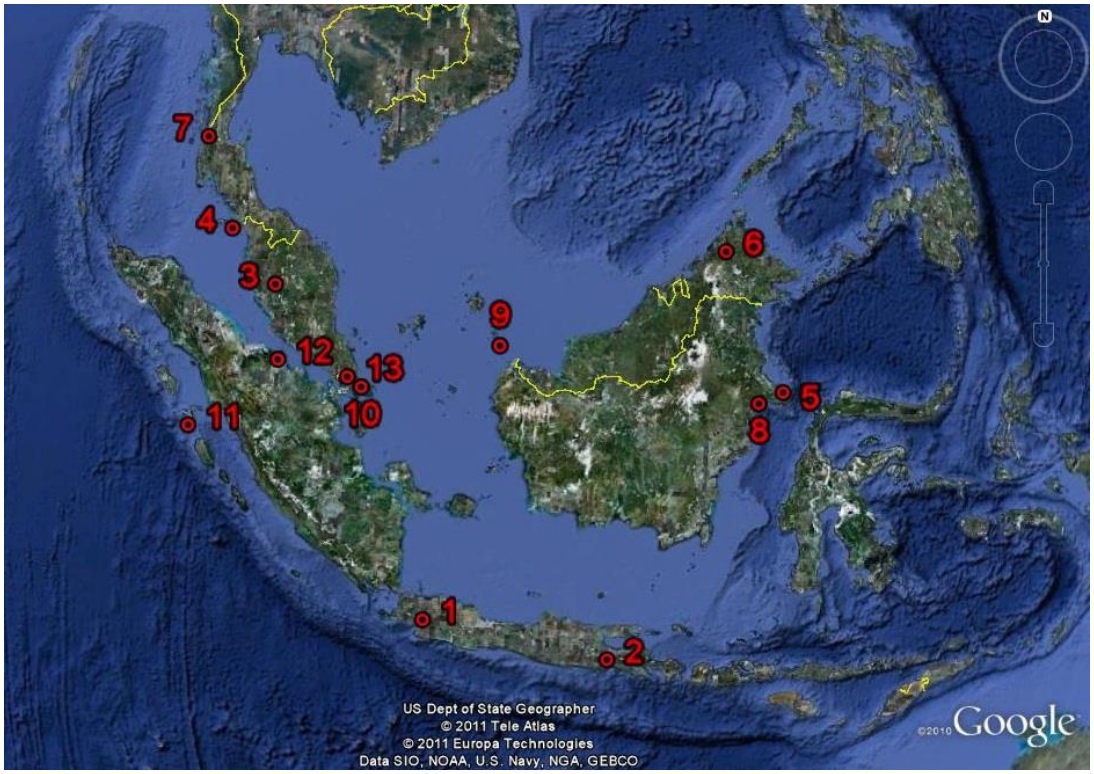


Figure A2.1. Geographic origin of USNM colugo specimens sampled.

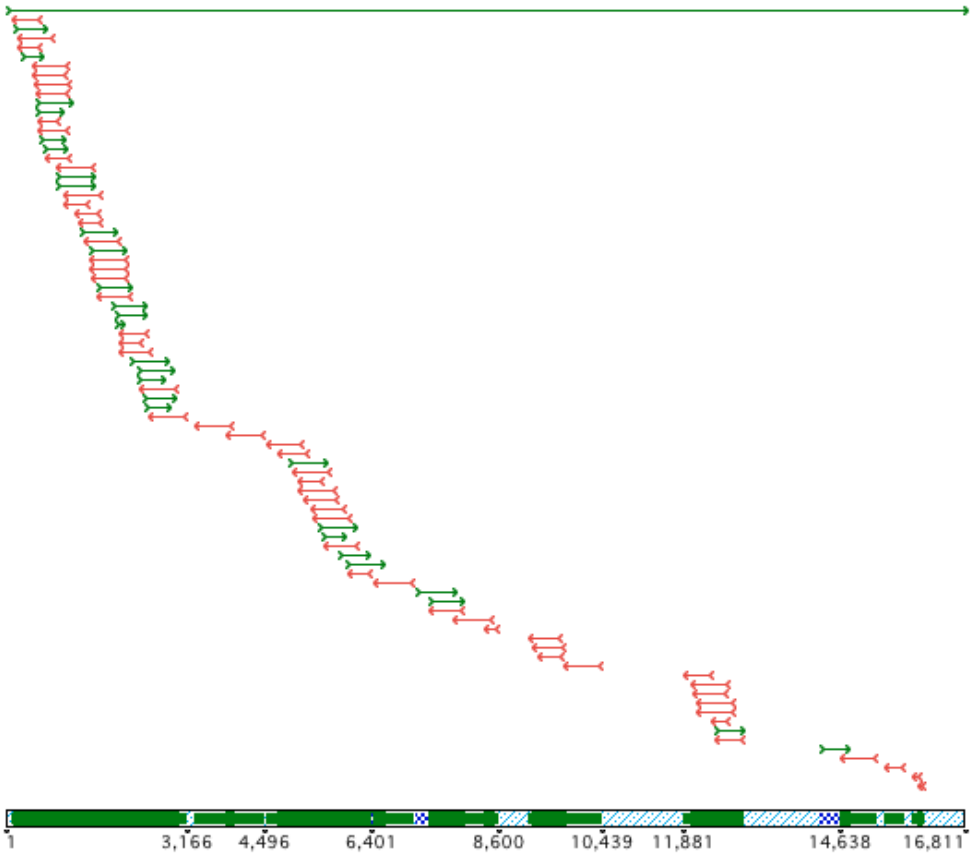


Figure A2.2. Mitochondrial genome coverage based on preliminary Sanger sequencing of 96 clones from specimens a) 12, and b). 6 relative to the reference sequence of a Bornean colugo (GenBank accession number AJ428849) represented by the long green arrow at the top of the figure. The smaller sequences (short red and green arrows) are random fragments recovered after 2 rounds of mtDNA selection that were cloned into a plasmid vector. 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 dark 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. A) Distribution of reads from specimen 12 mapped onto the colugo reference genome. B) Distribution of reads from specimen 6 mapped onto the colugo reference genome

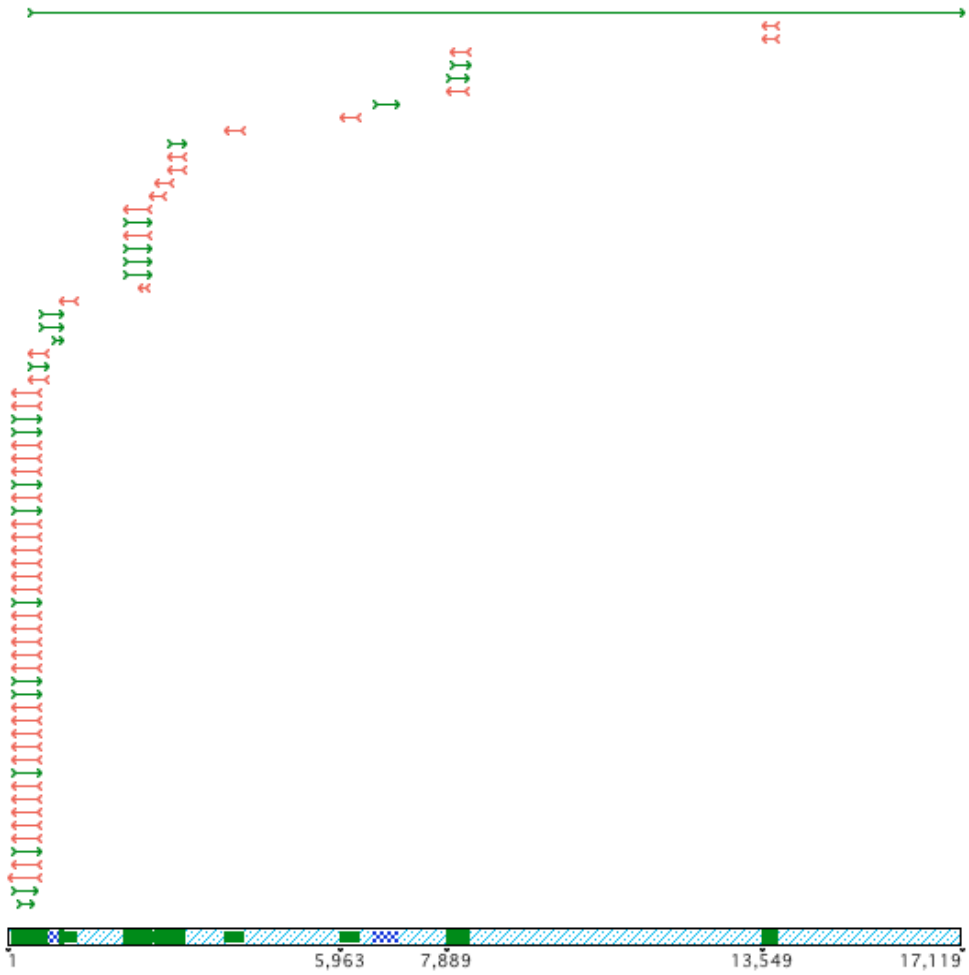


Figure A2.2 Continued.

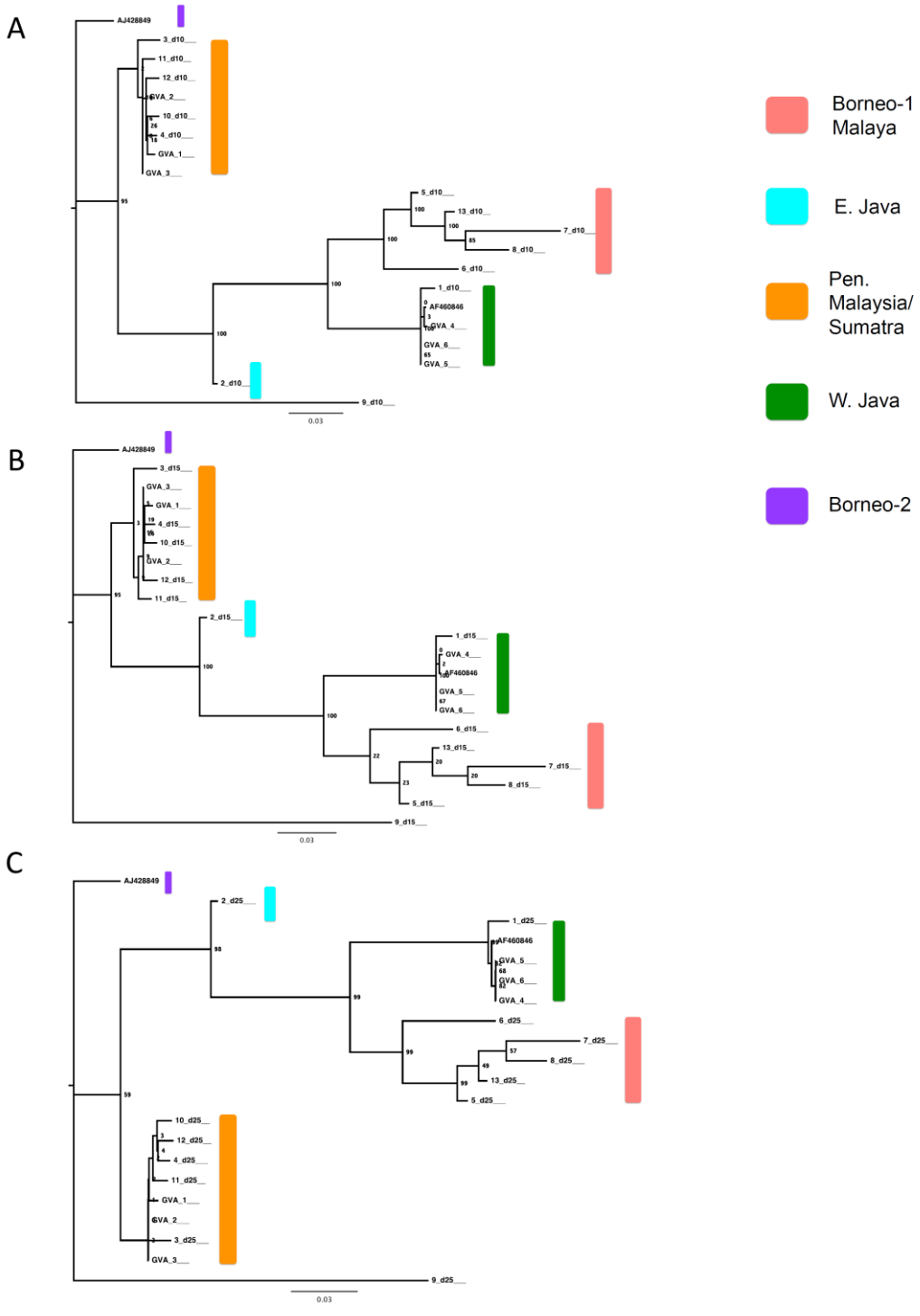


Figure A2.3. Maximum likelihood phylogenetic trees at variable sequence depth (d) enforcements. A) ML tree constructed from alignment d10. B) ML tree constructed from alignment d15. C) ML tree constructed from alignment d25.

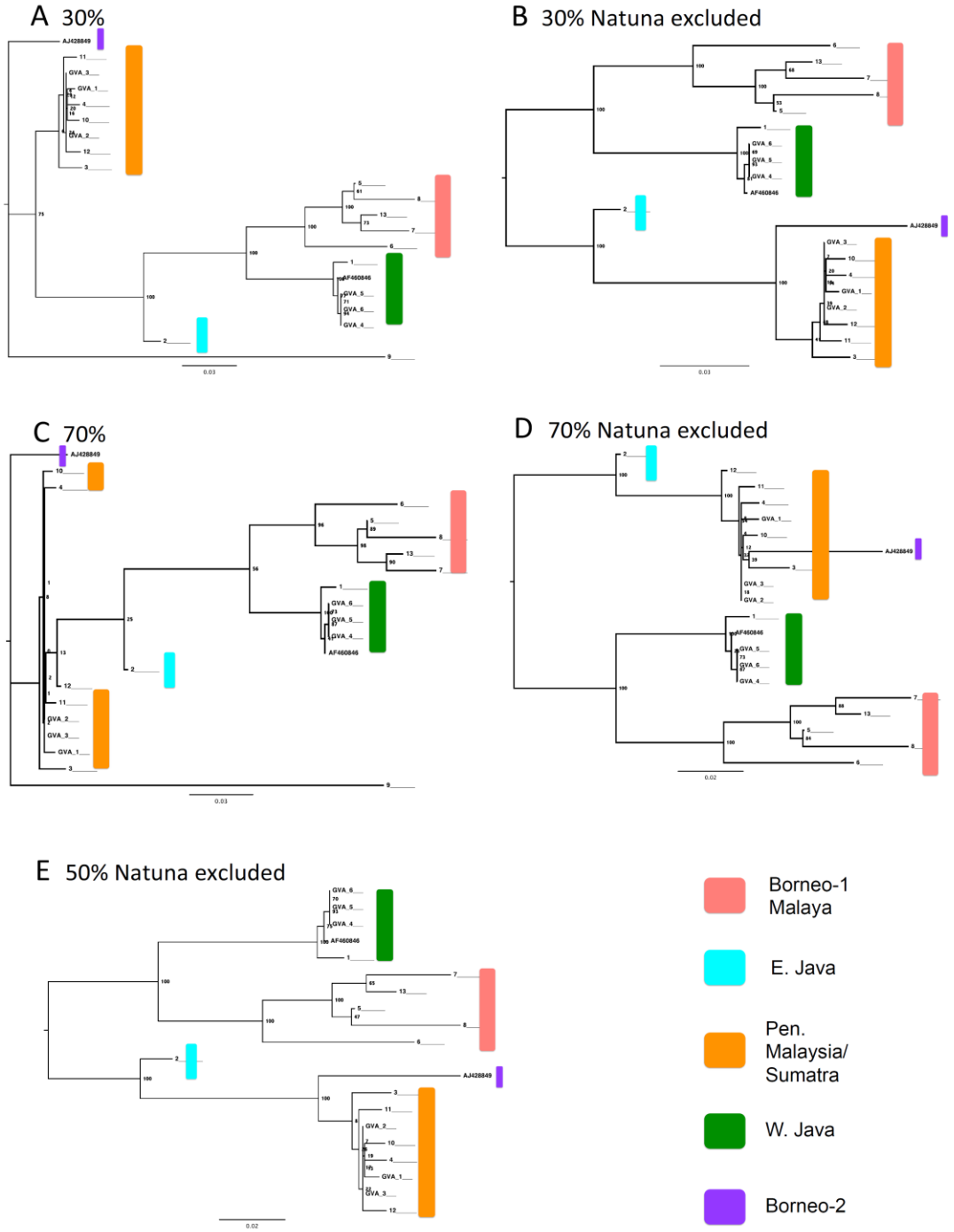


Figure A2.4. Maximum likelihood phylogenetic trees constructed from alignments where each site is present 30%, 50%, and 70% of sites across individuals. Analyses (B, D, E) are shown where the Natuna Islands specimen (9) was removed to examine effects of long-branch attraction on the phylogenetic support levels. The 50% tree is shown in Fig. 5, while the Natuna (9)-excluded tree for 50% is shown in E.

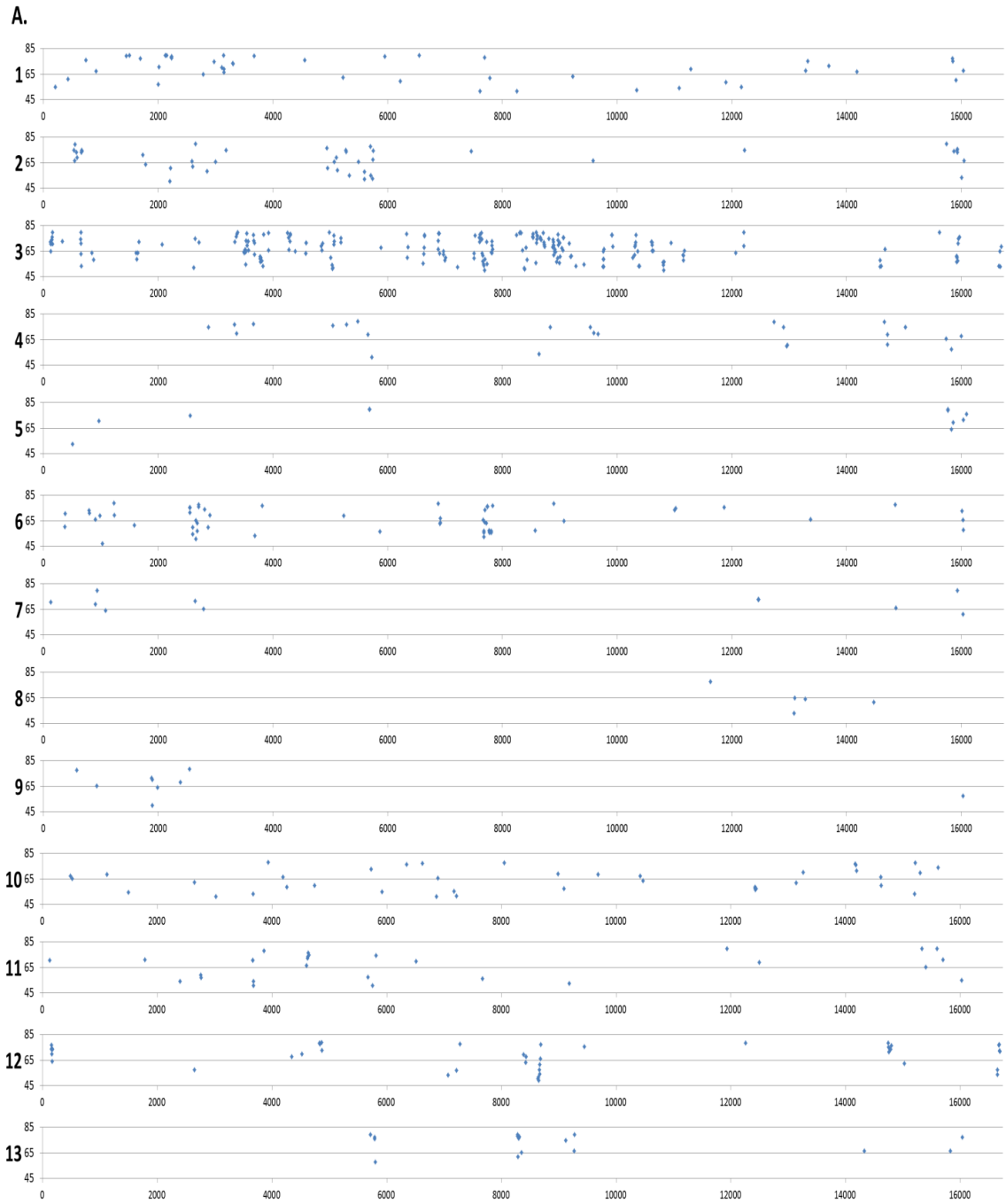


Figure A2.5. Biallelic SNP locations for each individual and SNP statistics. Only SNPs with minor allele frequency ≥20% were included. A) Major allele frequencies are plotted in their relative position across the mitochondrial genome for each individual. B) SNP Statistics.

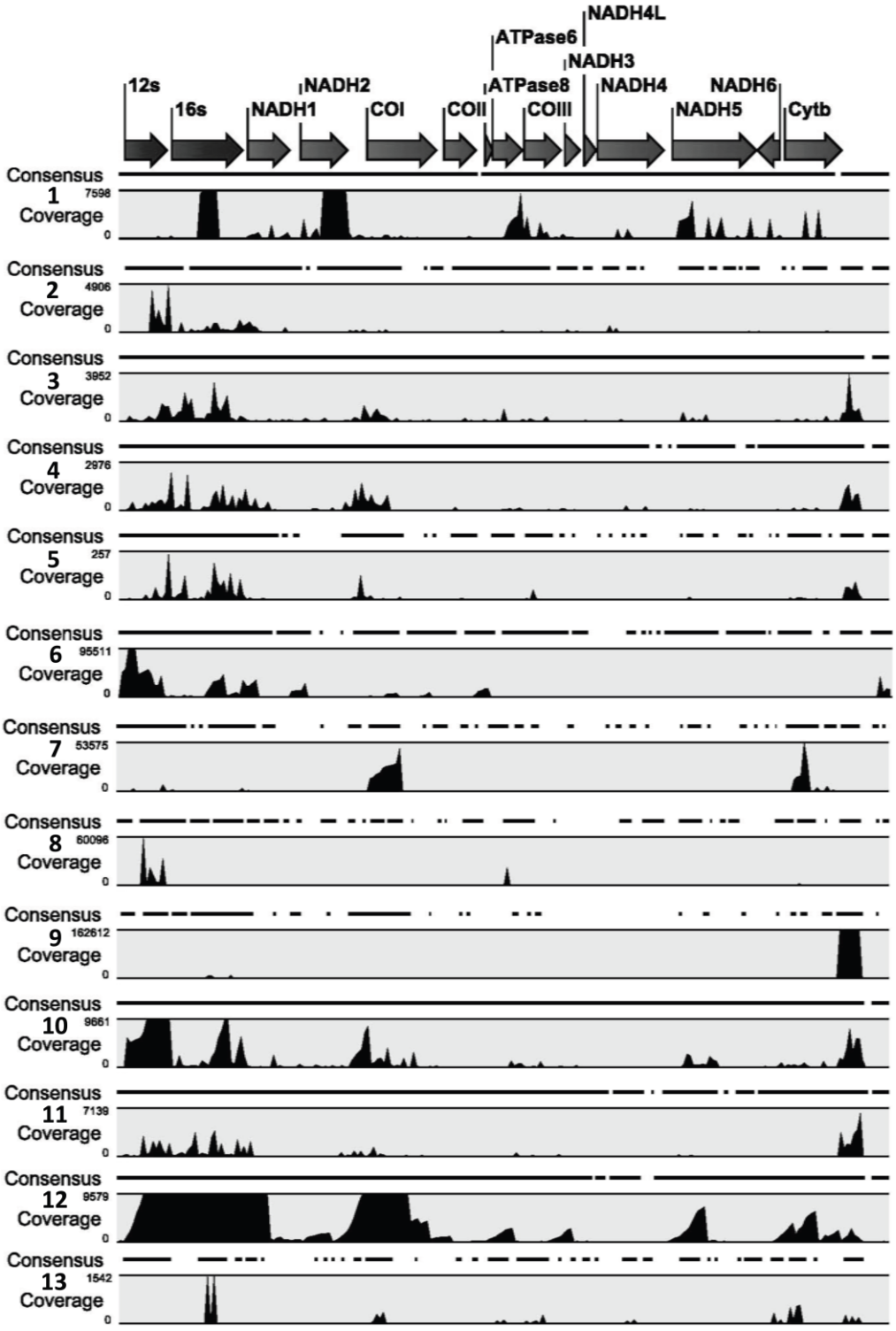


Figure A2.6. Unabbreviated sequence depth maps of each individual.

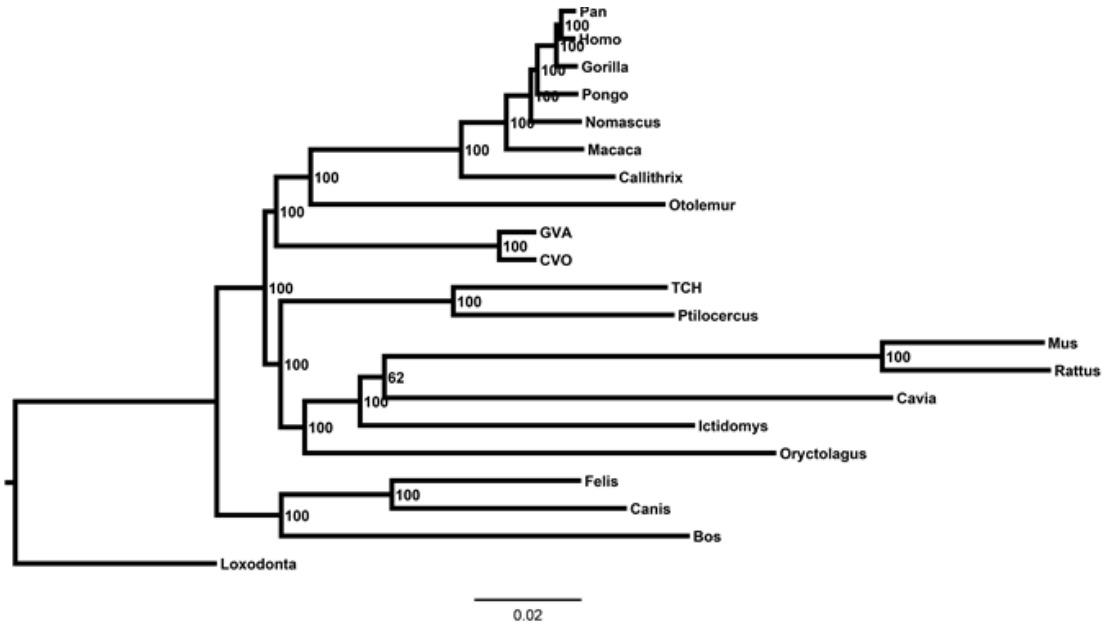


Figure A3.1. Maximum likelihood tree based on a concatenated alignment of 634 1:1 orthologous CDS, using all codon positions. Bootstrap values are based on 1000 pseudoreplicates. Identical results were obtained when the analysis was restricted to first, second, or first+second codon positions, or amino acid sequences.

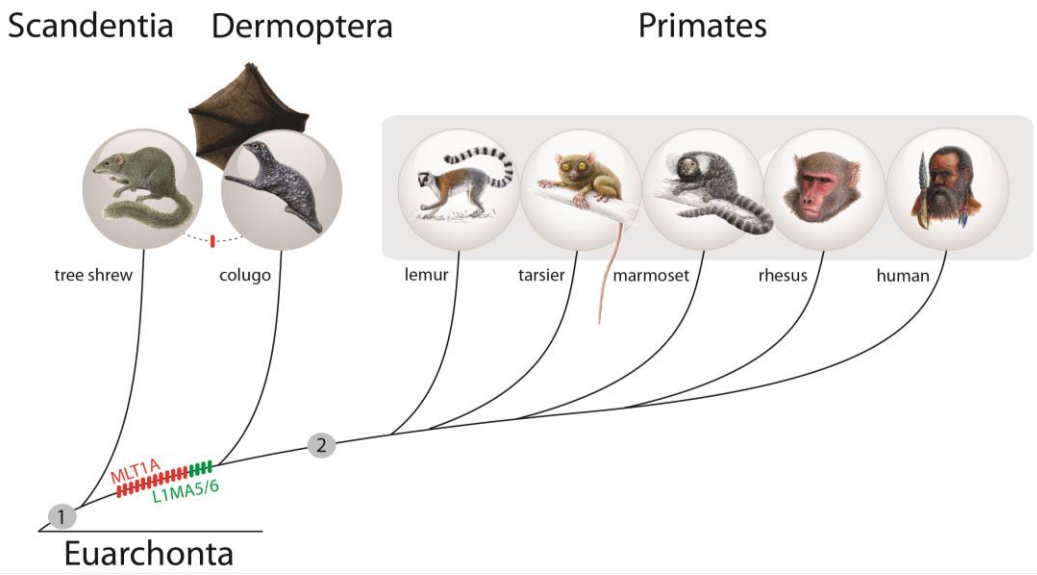


Figure A3.2. Phylogenetic tree of Euarchonta. 1) The monophyly of Euarchonta (Scandentia, Dermoptera, and Primates) is supported by one MLT1A and four MSTD elements (*11*), and 2) the monophyly of primates is confirmed by synapomorphic retrotransposon insertions (*6*). The close relationship of primates and colugo is now supported by 12 MLT1A (red) and 4 L1MA5/6 (green) elements. One MLT1A insertion supports a sister-group relationship of colugo and tree shrew (at the dotted line between the two). This apparently conflicting pattern is likely the result of incomplete lineage sorting.

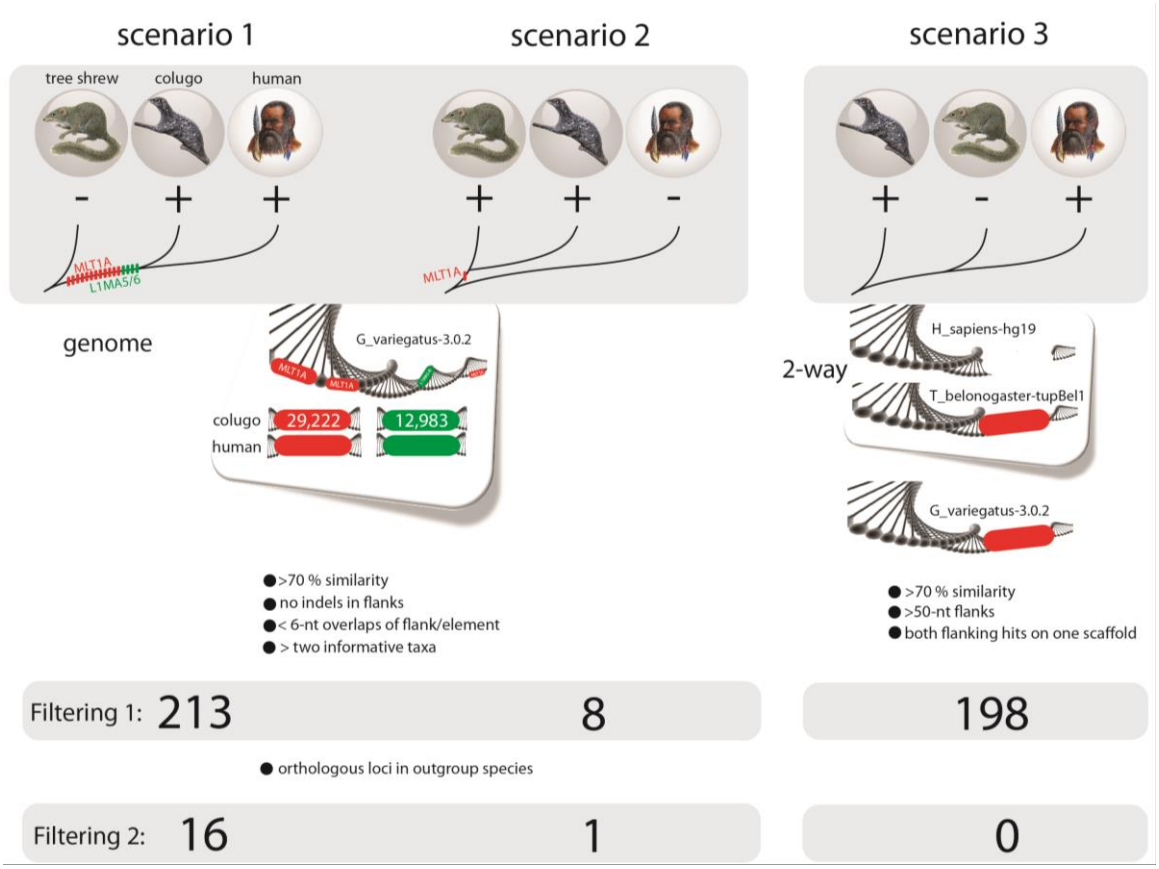


Figure A3.3. Description of the retroposon insertion screening strategy. The three possible phylogenetic scenarios for the relatedness of colugos are presented at the top. Scenario 1 describes a close relationship between the colugo and primates, the second places the tree shrew and colugo on one branch, and scenario 3 proposes a close relationship between the tree shrew and human.

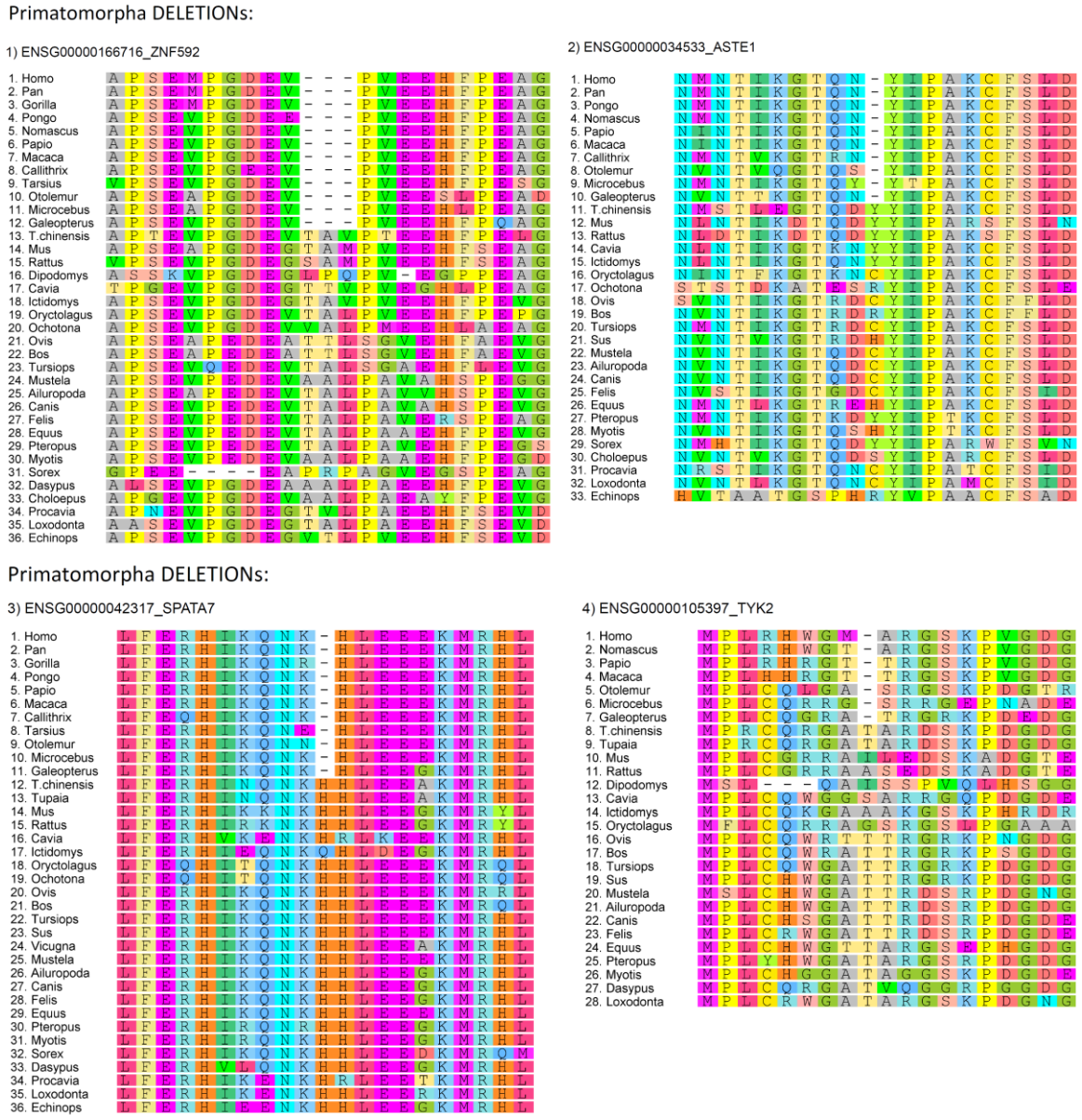


Figure A3.4. Phylogenetically informative indels supporting (P1-P15) Primatomorpha, (S1-S5) Sundatheria, and (PS1-PS5) Primates + Scandentia. (*Following pages*).

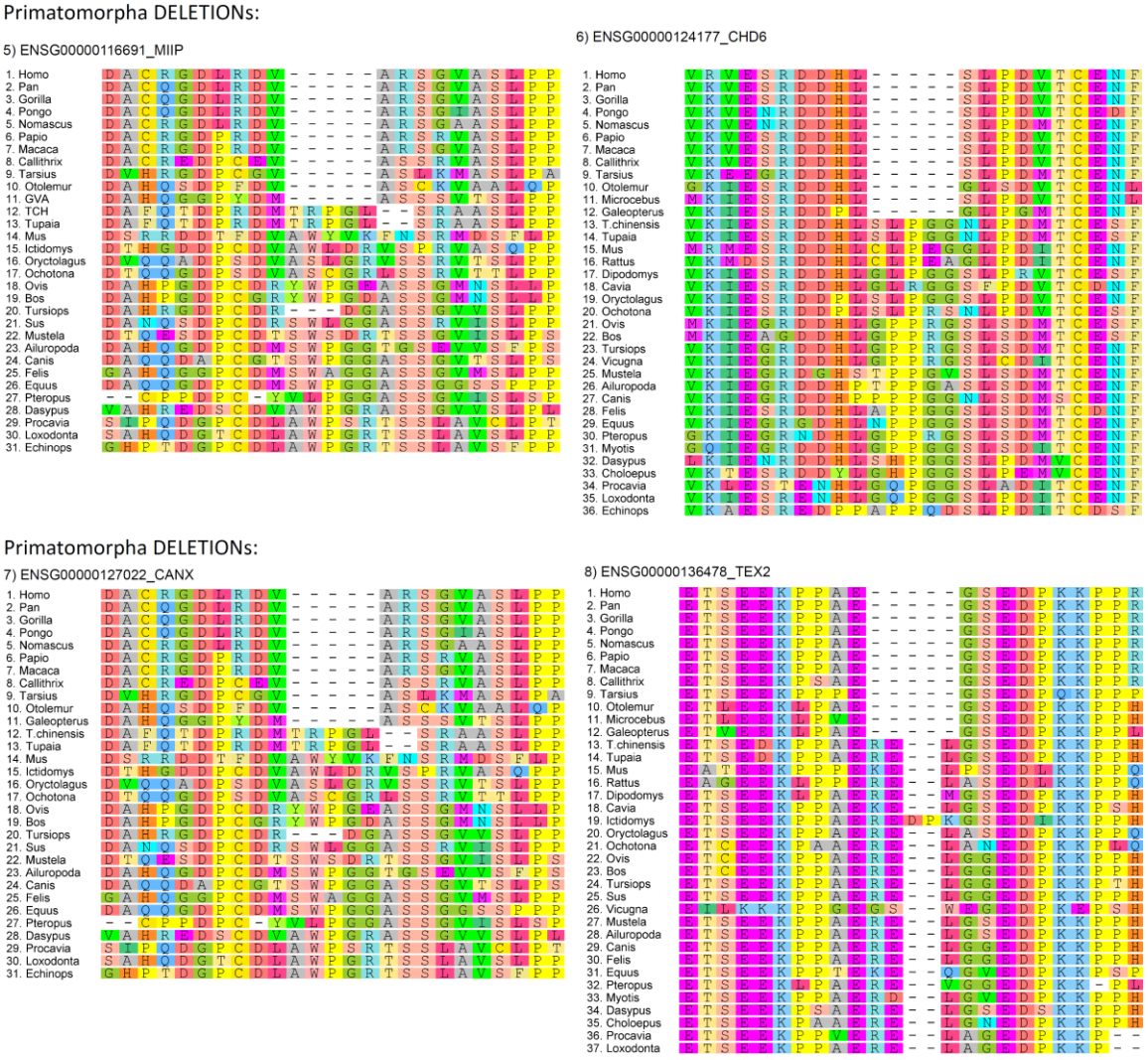


Figure A3.4 Continued.

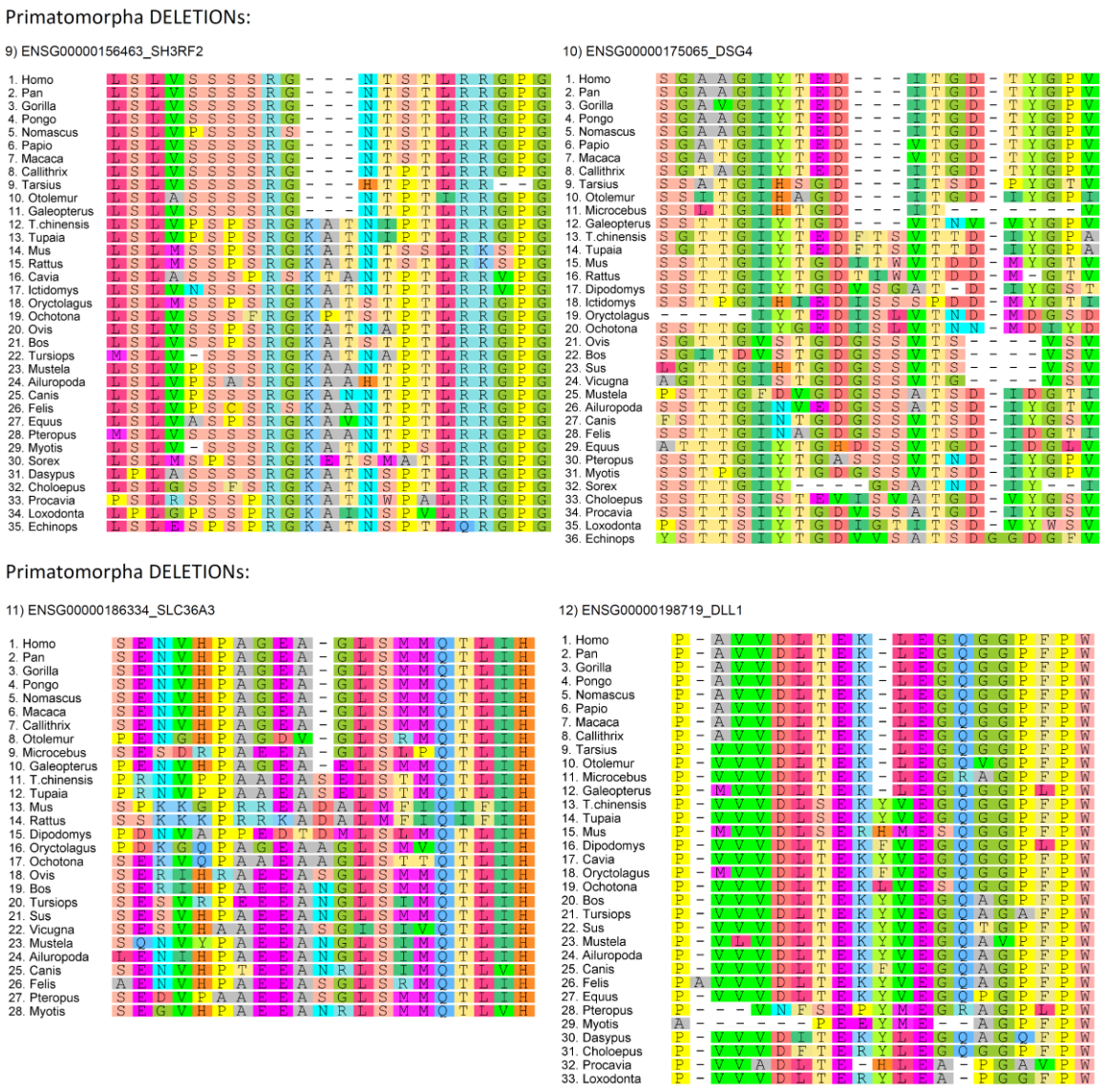
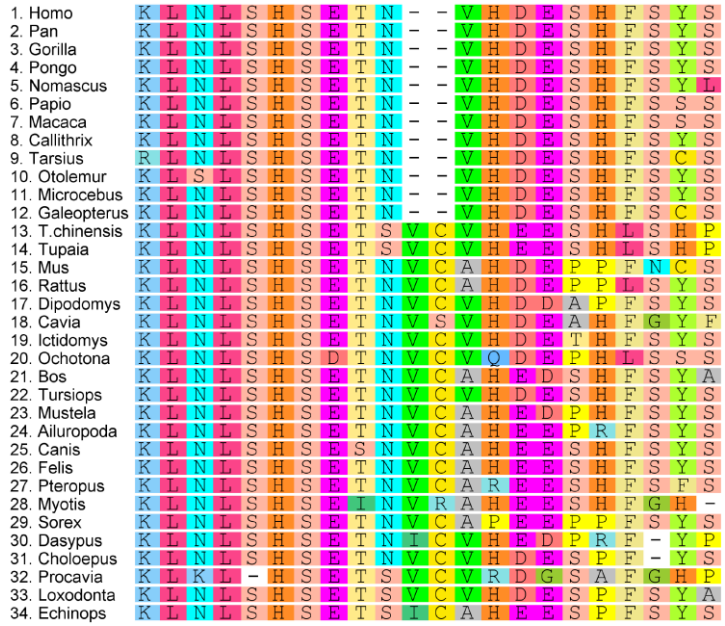


Figure A3.4 Continued.

Primatomorpha DELETIONs:

13) ENSG00000275342_NONAME



Primatomorpha INSERTION:

14) ENSG00000079691_LRRC16A

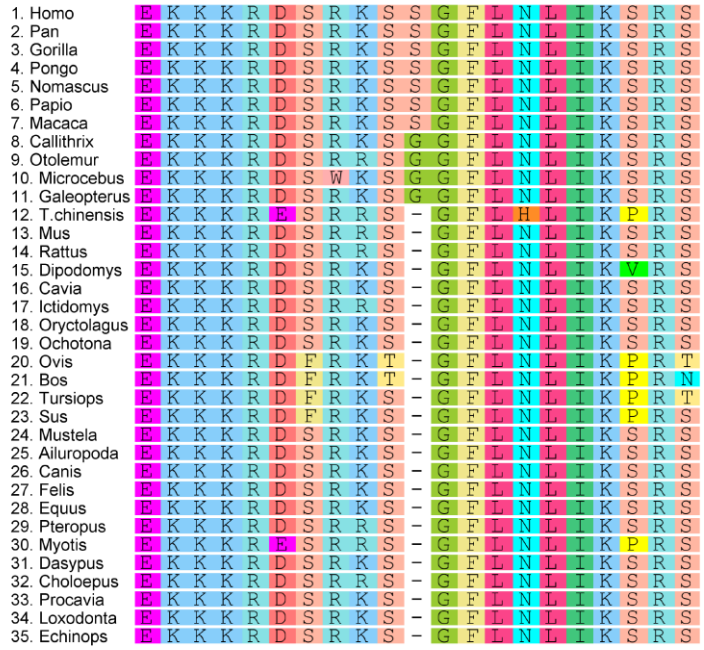


Figure A3.4 Continued.

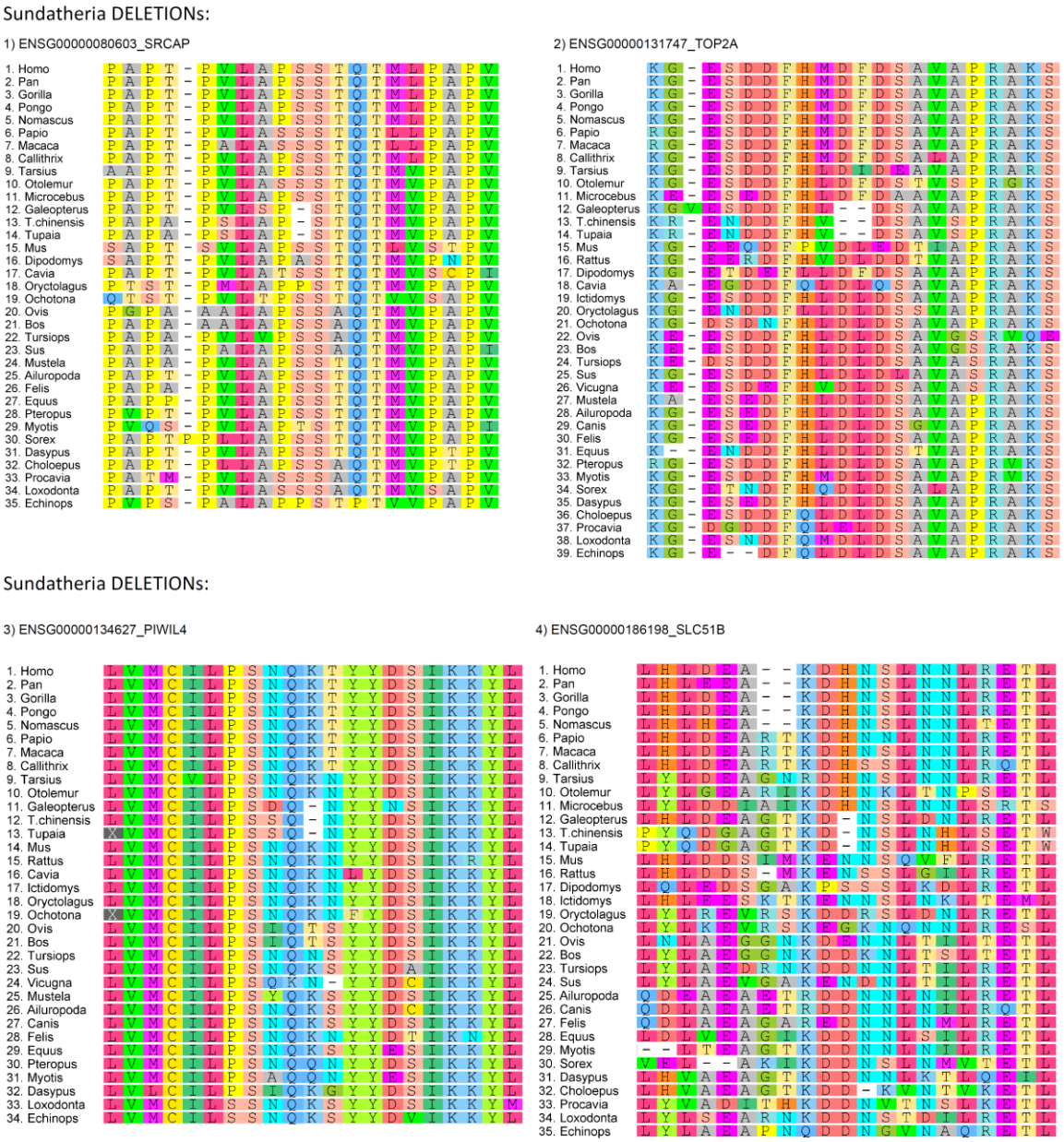


Figure A3.4 Continued.

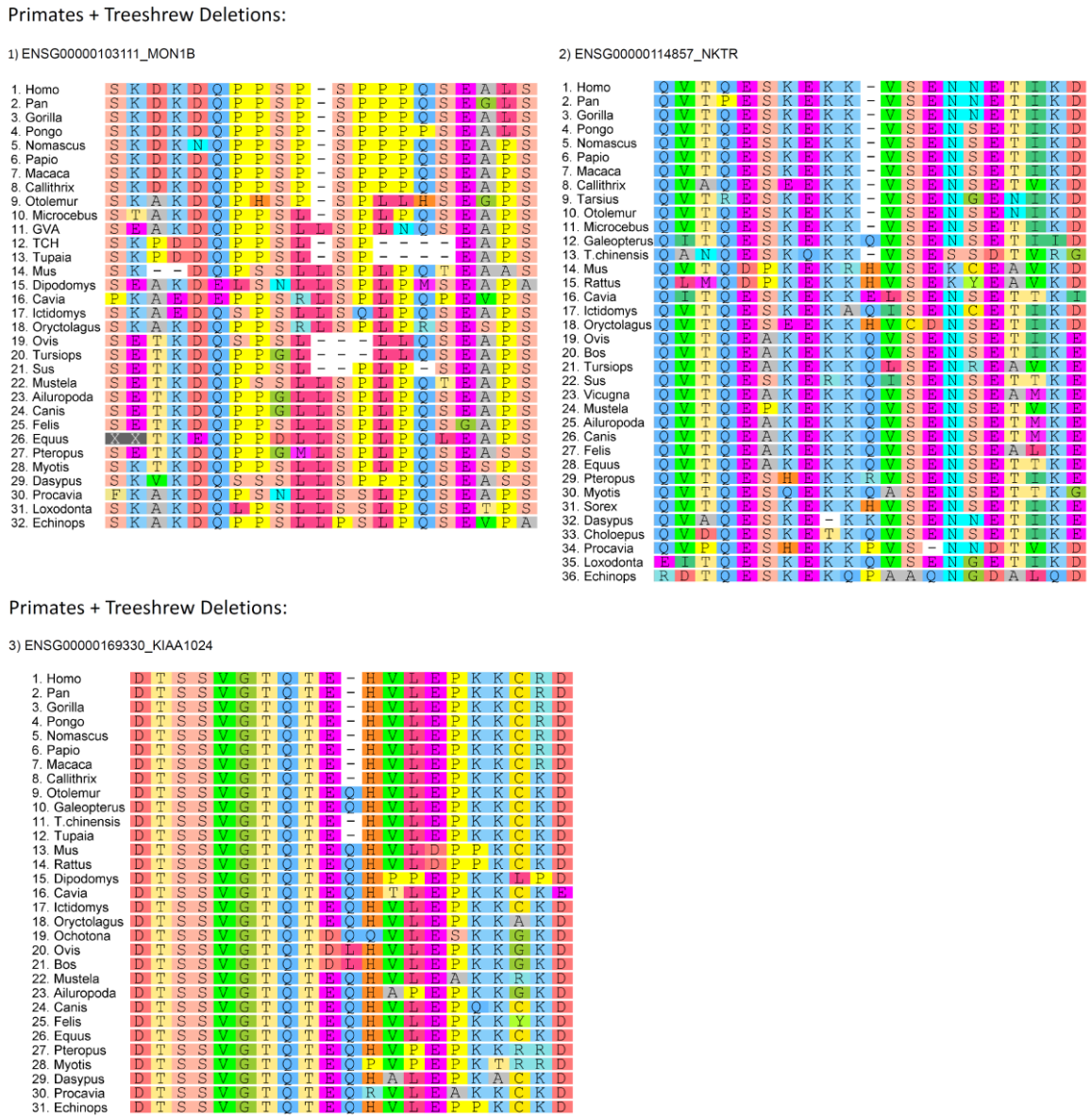


Figure A3.4 Continued.

Primates + Treeshrew Insertions:

4) ENSG00000253309_SERPINE3

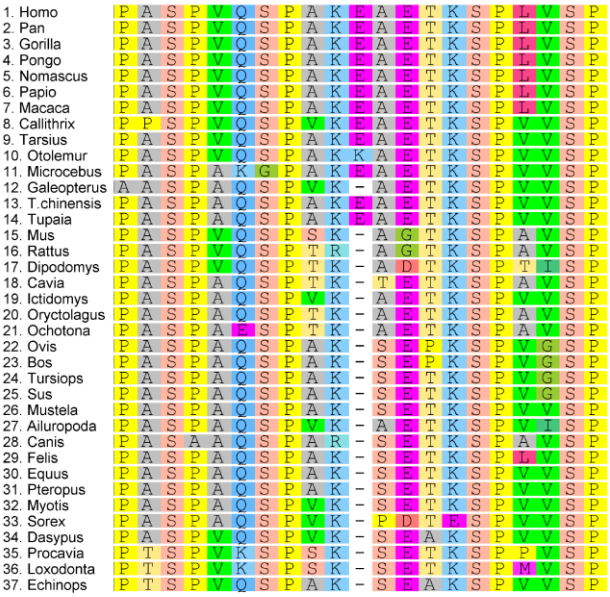


Figure A3.4 Continued.

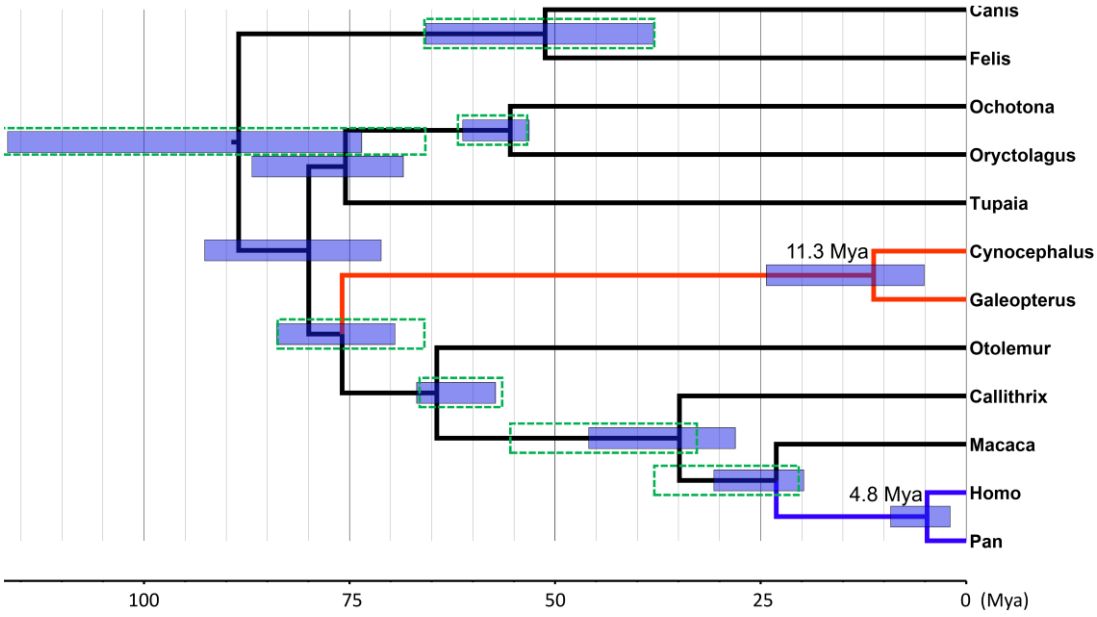


Figure A3.5. Time-tree showing seven external fossil calibration ranges as green boxes (Table B3.7). Topology was based on the maximum likelihood 'best-tree'. The human/chimpanzee bifurcation (blue) node was left uncalibrated to serve as a control for terminal node age estimates that lack calibrations. Blue boxes are calculated 95% confidence intervals.

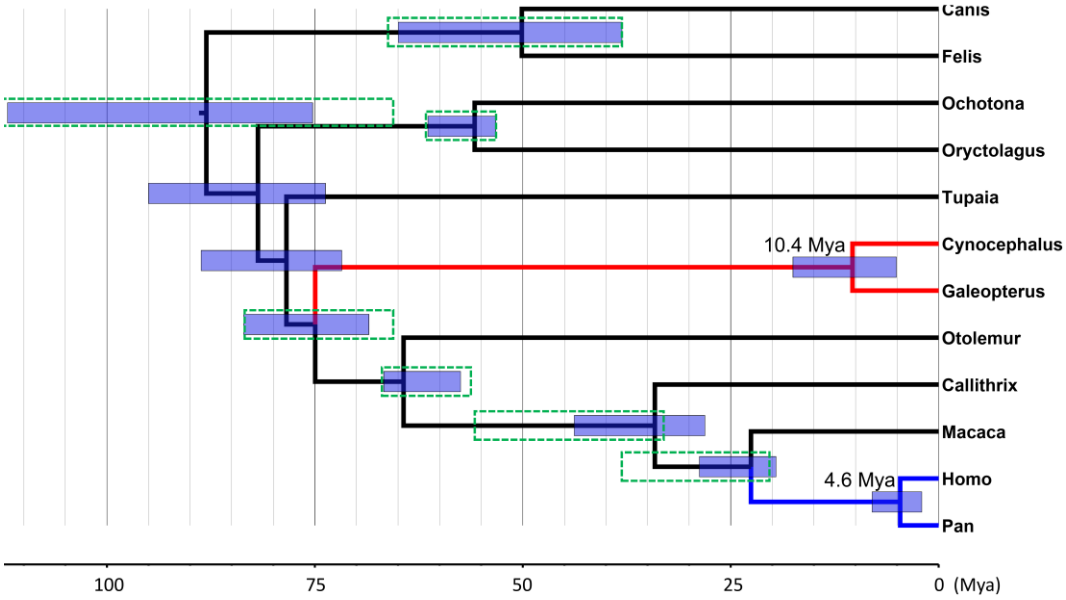


Figure A3.6. Time-tree with showing seven external fossil calibrations as green boxes (Table B3.7). Euarchontan monophyly was enforced on the maximum likelihood 'best-tree' topology. The human/chimpanzee node (blue) was left to serve as a control for terminal node age estimates that lack calibrations. Blue boxes are calculated 95% confidence intervals.

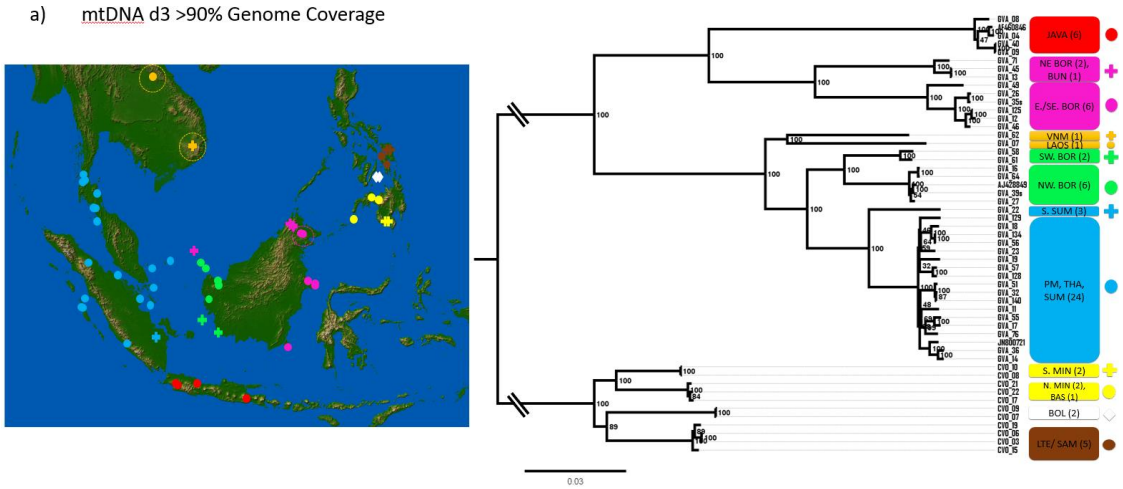


Figure A3.7. Maximum likelihood mtDNA tree with 1000 bootstrap iterations, 53 taxa, and greater than 90% mitogenome coverage. Labels for geographic area are next to sample numbers. Abbreviations: Vietnam=VNM, Thailand=THA, Peninsular Malaysia=PM, Sumatra=SUM, Borneo=BOR, Pulau Bunguran=BUN, Pulau Basilan=BAS, Mindanao=MIN, Bohol=BOL, Leyte=LTE, Samar=SAM, Pulau Dinagat=DIN.

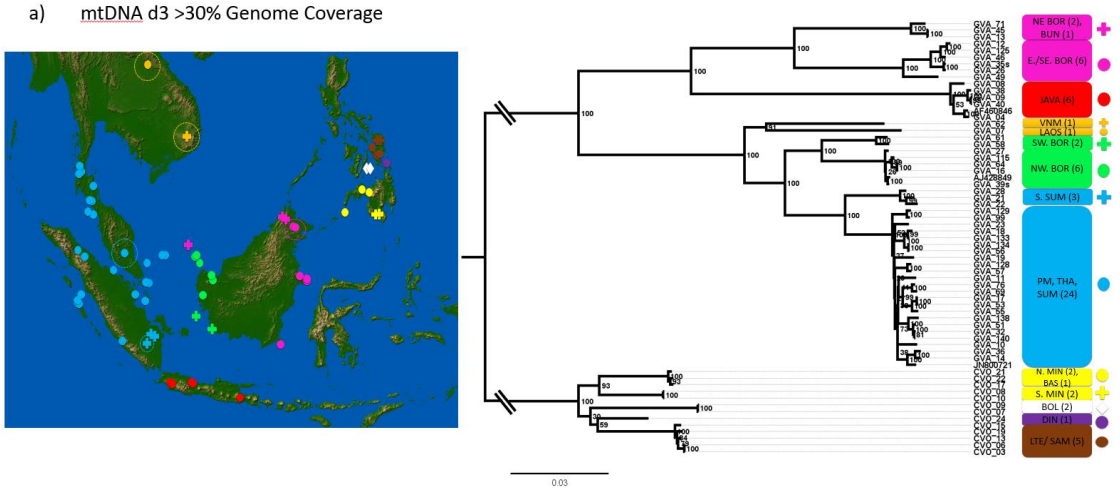


Figure A3.8. Maximum likelihood mtDNA tree with 1000 bootstrap iterations 65 taxa, and greater than 30% mitogenome coverage. Labels for geographic area are next to sample numbers. Abbreviations: Vietnam=VNM, Thailand=THA, Peninsular Malaysia=PM, Sumatra=SUM, Borneo=BOR, Pulau Bunguran=BUN, Pulau Basilan=BAS, Mindanao=MIN, Bohol=BOL, Leyte=LTE, Samar=SAM, Pulau Dinagat=DIN.

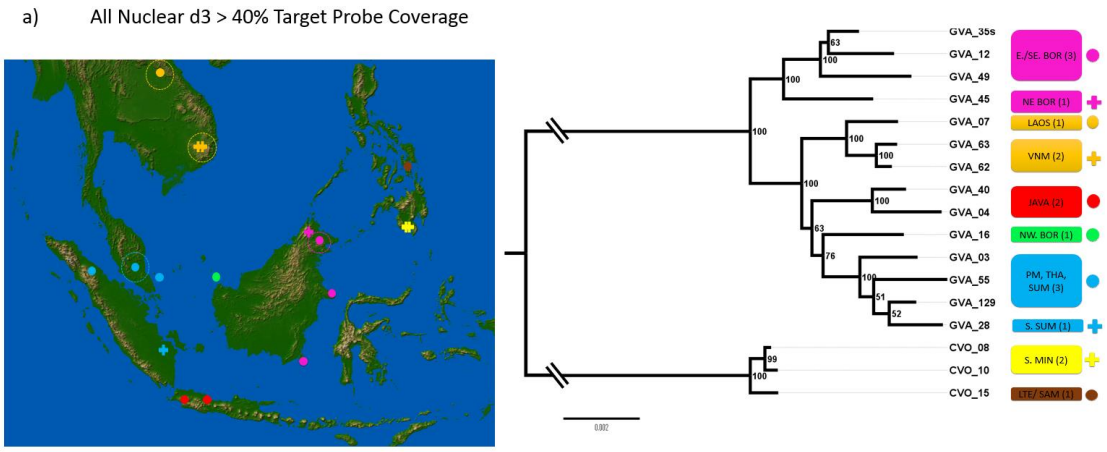


Figure A3.9. Maximum likelihood all-nuclear (biparental+Chr Y) tree with 1000 bootstrap iterations with sequence depth ≥3. Labels for geographic area are next to sample numbers. Abbreviations: Vietnam=VNM, Thailand=THA, Peninsular Malaysia=PM, Sumatra=SUM, Borneo=BOR, Pulau Bunguran=BUN, Pulau Basilan=BAS, Mindanao=MIN, Bohol=BOL, Leyte=LTE, Samar=SAM, Pulau Dinagat=DIN.

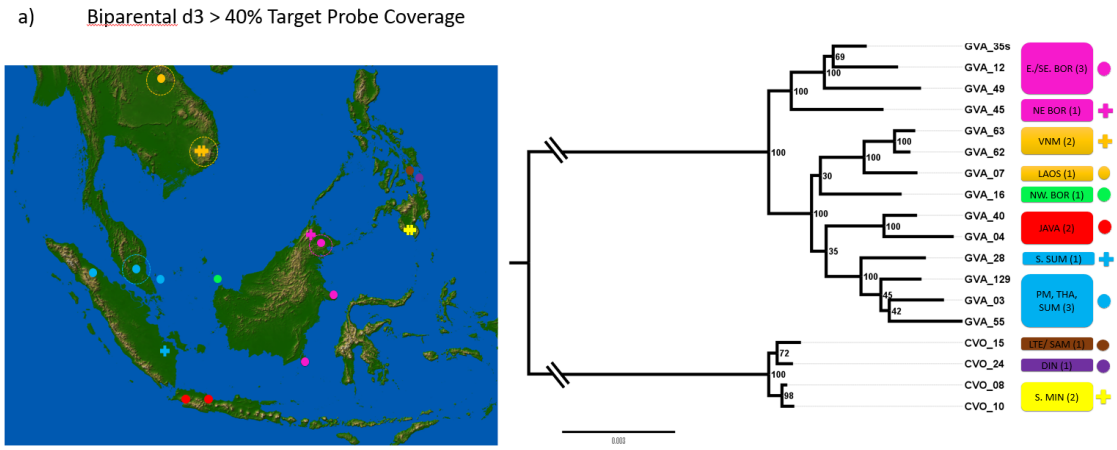


Figure A3.10. Maximum likelihood biparental locus tree with 1000 bootstrap iterations with minimum sequence depth ≥2. Abbreviations: Vietnam=VNM, Thailand=THA, Peninsular Malaysia=PM, Sumatra=SUM, Borneo=BOR, Pulau Bunguran=BUN, Pulau Basilan=BAS, Mindanao=MIN, Bohol=BOL, Leyte=LTE, Samar=SAM, Pulau Dinagat=DIN.

a) Y-Chromosome d2 >15% Probe Coverage

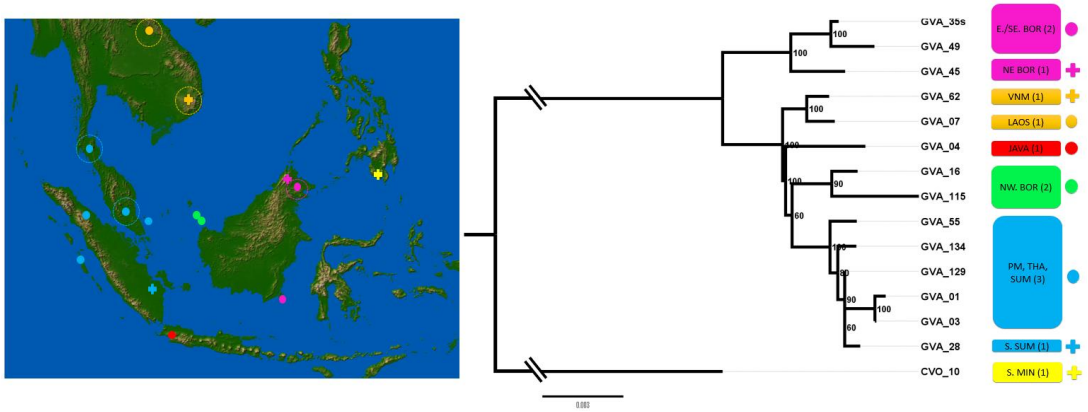


Figure A3.11. Maximum likelihood Y-chromosome tree (minimum depth ≥2) with 1000 bootstrap iterations. Abbreviations: Vietnam=VNM, Thailand=THA, Peninsular Malaysia=PM, Sumatra=SUM, Borneo=BOR, Pulau Bunguran=BUN, Pulau Basilan=BAS, Mindanao=MIN, Bohol=BOL, Leyte=LTE, Samar=SAM, Pulau Dinagat=DIN.

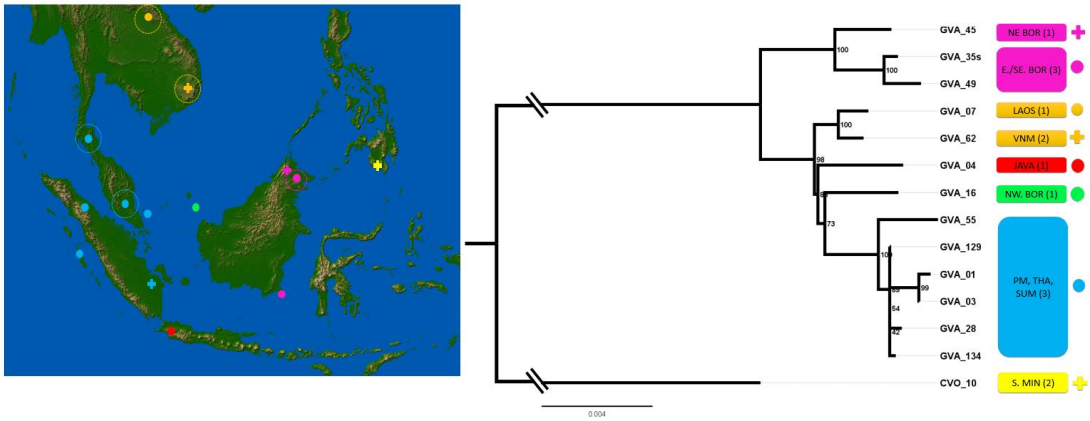


Figure A3.12. Maximum likelihood Y-chromosome tree (minimum depth $\geq$ 3) with 1000 bootstrap iterations. Abbreviations: Vietnam=VNM, Thailand=THA, Peninsular Malaysia=PM, Sumatra=SUM, Borneo=BOR, Pulau Bunguran=BUN, Pulau Basilan=BAS, Mindanao=MIN, Bohol=BOL, Leyte=LTE, Samar=SAM, Pulau Dinagat=DIN.

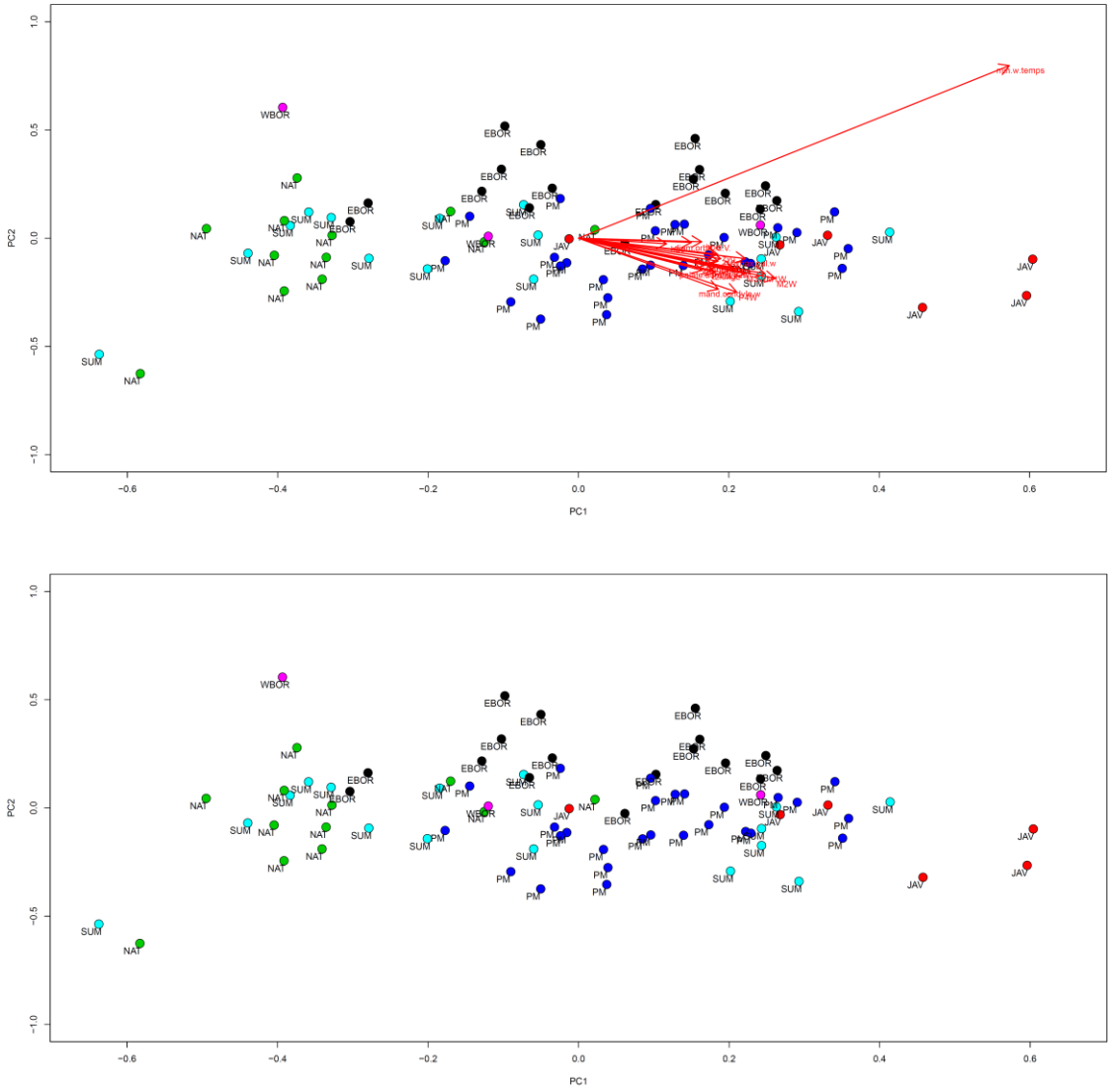


Figure A3.13. PCA analysis (singular value decomposition) of 19 craniodental measurments for male and female Sunda colugos. Dwarf individuals were not included. Red arrows (top) are the 19 variable vectors. Each point represents one individual. 45% of the variation is explained by PC1 and 29% by PC2.

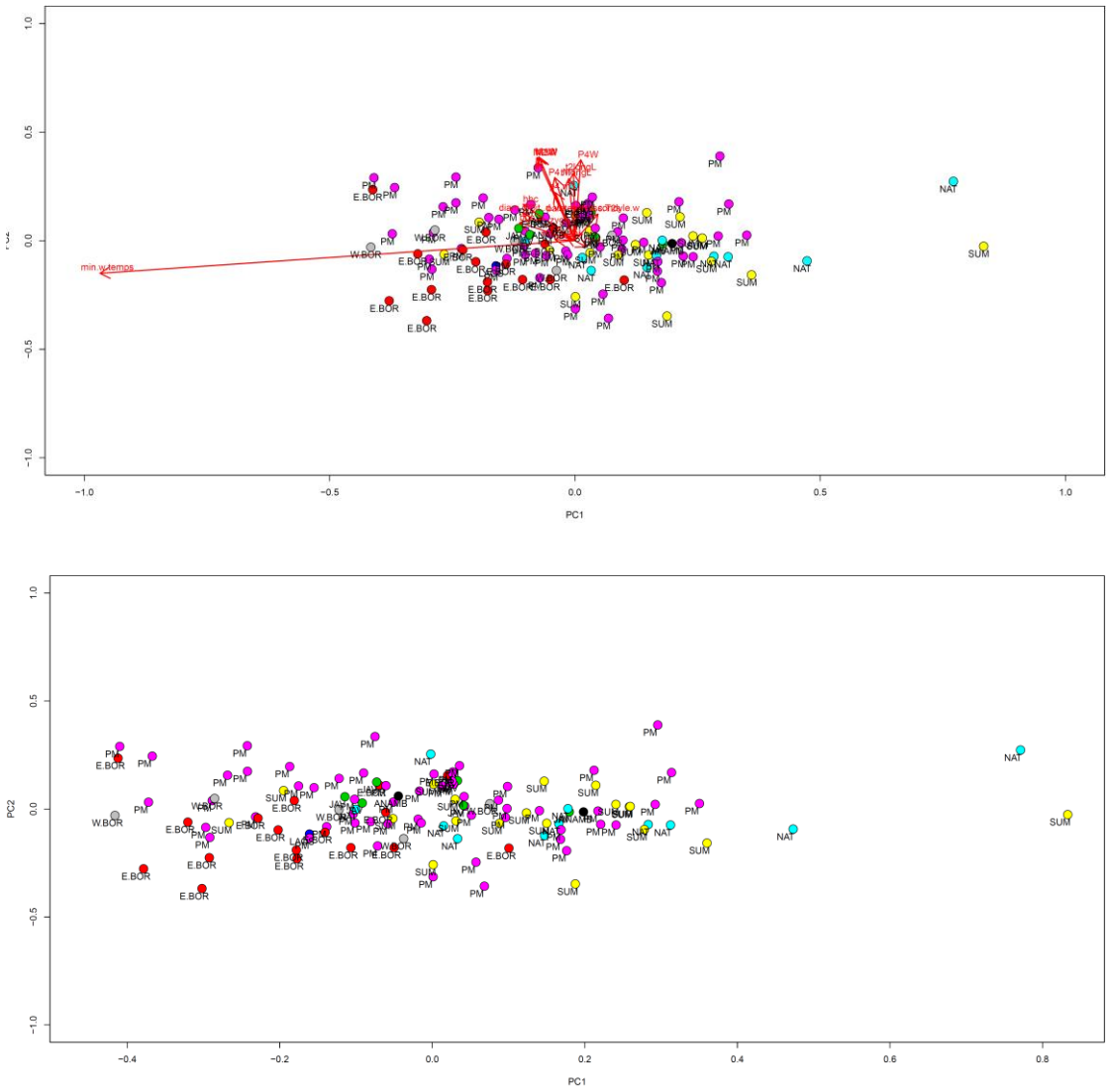


Figure A3.14. PCA analysis (singular value decomposition) of 19 craniodental measurments for male and female Sunda colugos, with dwarf individuals included for comparison to Fig. S18. Red arrows are variable vectors. Each point is an individual. All of the craniodental measurements are normalized for body size. 44% of the variation is explained by PC1 and 20% by PC2.

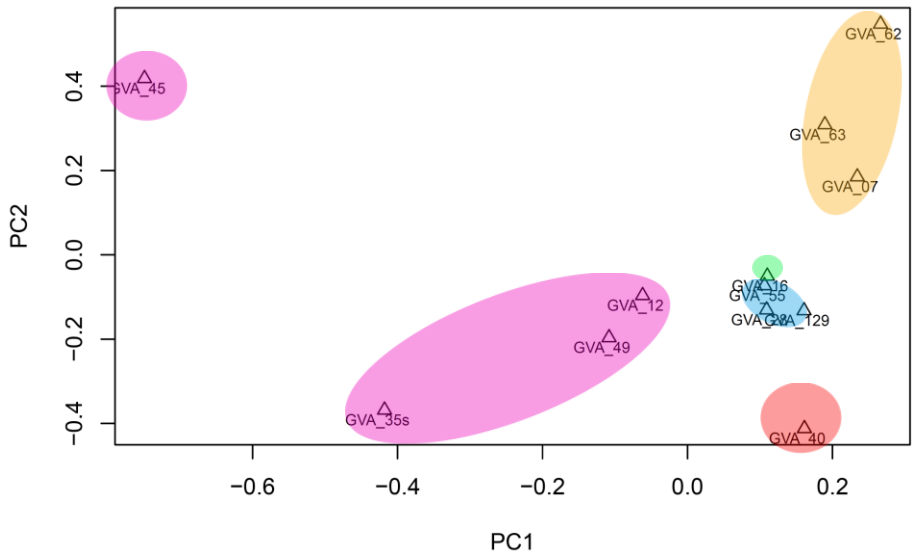


Figure A3.15. Principal component analysis (PC1 vs. PC2) of X-chromosome variation in Sunda colugos. Colors follow geographic clusters in Fig. 3.

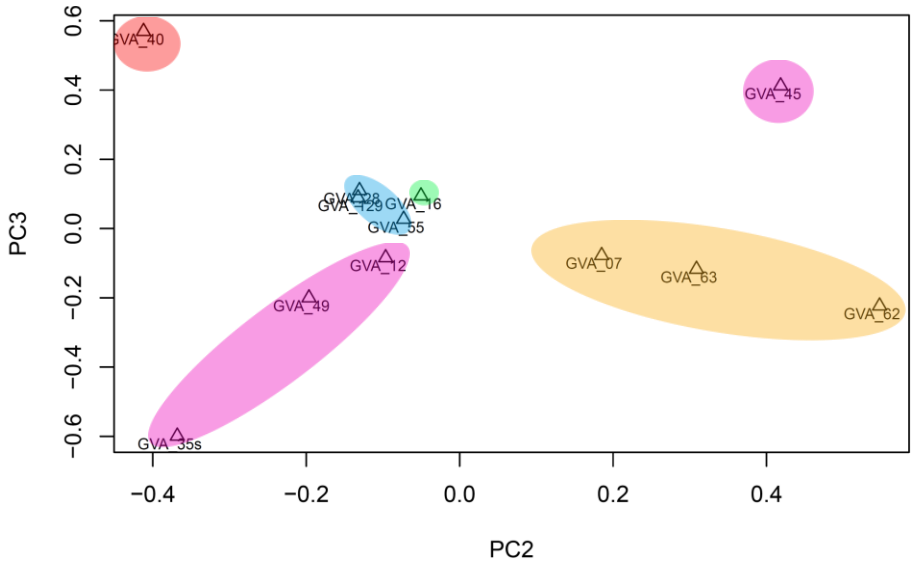


Figure A3.16. Principal component analysis (PC2 vs. PC3) of X-chromosome variation in Sunda colugos. Colors follow geographic clusters in Fig. 3.

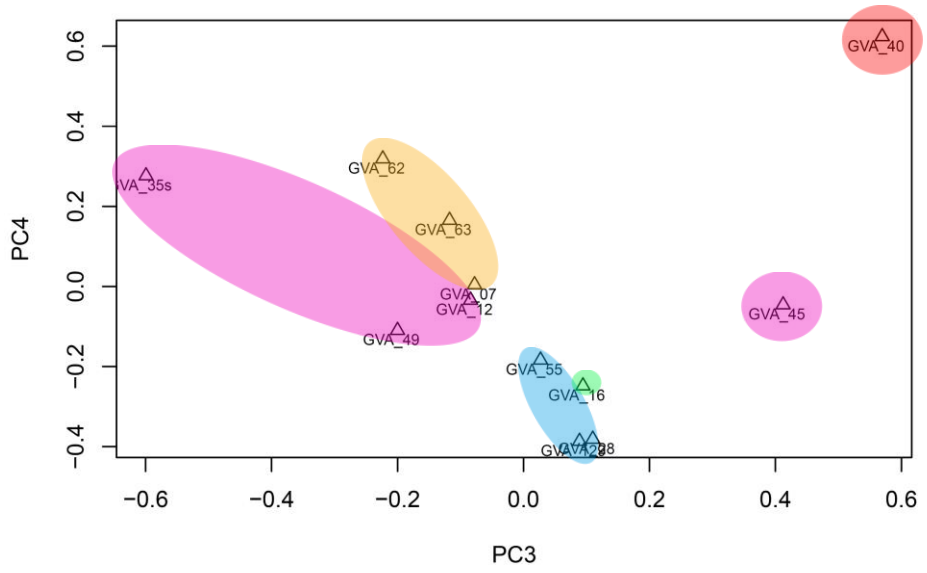


Figure A3.17. Principal component analysis (PC3 vs. PC4) of X-chromosome variation in Sunda colugos. Colors follow geographic clusters in Fig. 3.

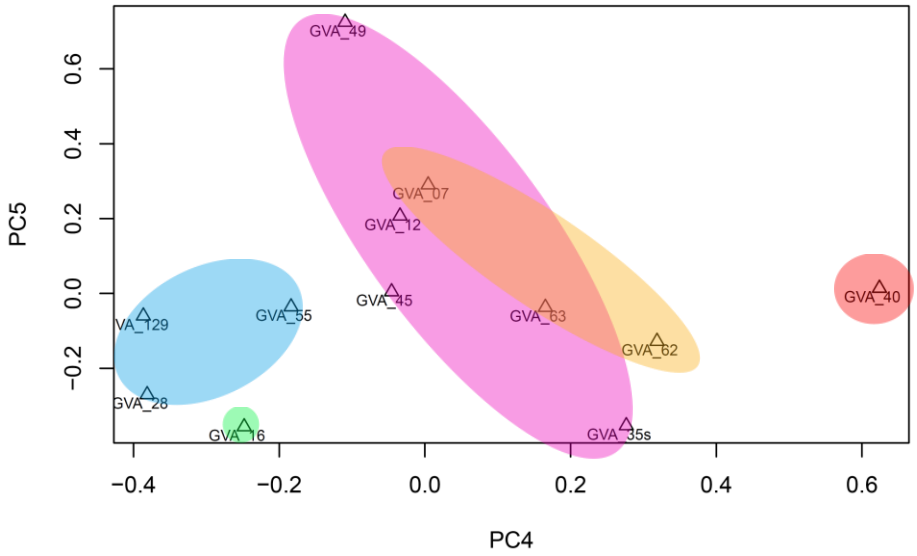


Figure A3.18. Principal component analysis (PC4 vs. PC5) of X-chromosome variation in Sunda colugos. Colors follow geographic clusters in Fig. 3.

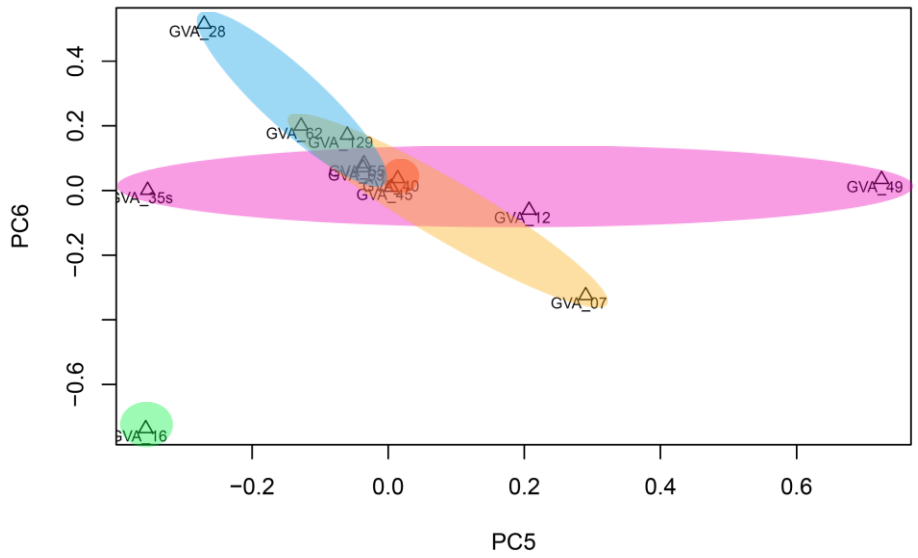


Figure A3.19. Principal component analysis (PC5 vs. PC6) of X-chromosome variation in Sunda colugos. Colors follow geographic clusters in Fig. 3. West Borneo (green) is well separated in this comparison.

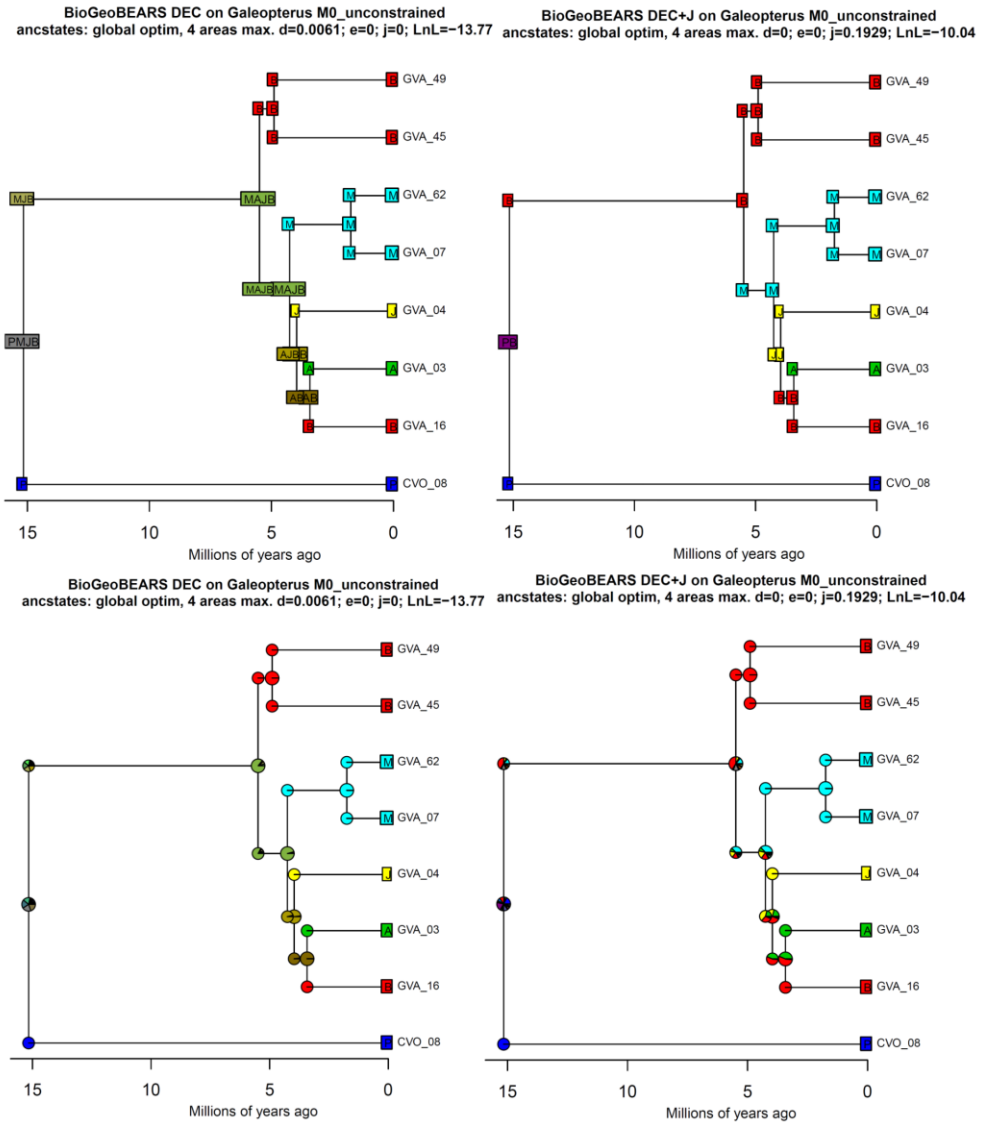


Figure A4.1. Ancenstral Area Probability calculations for the Sunda colugo with models DEC, DEC+j, DIVA, DIVA+j, BayArea, and BayArea+J.

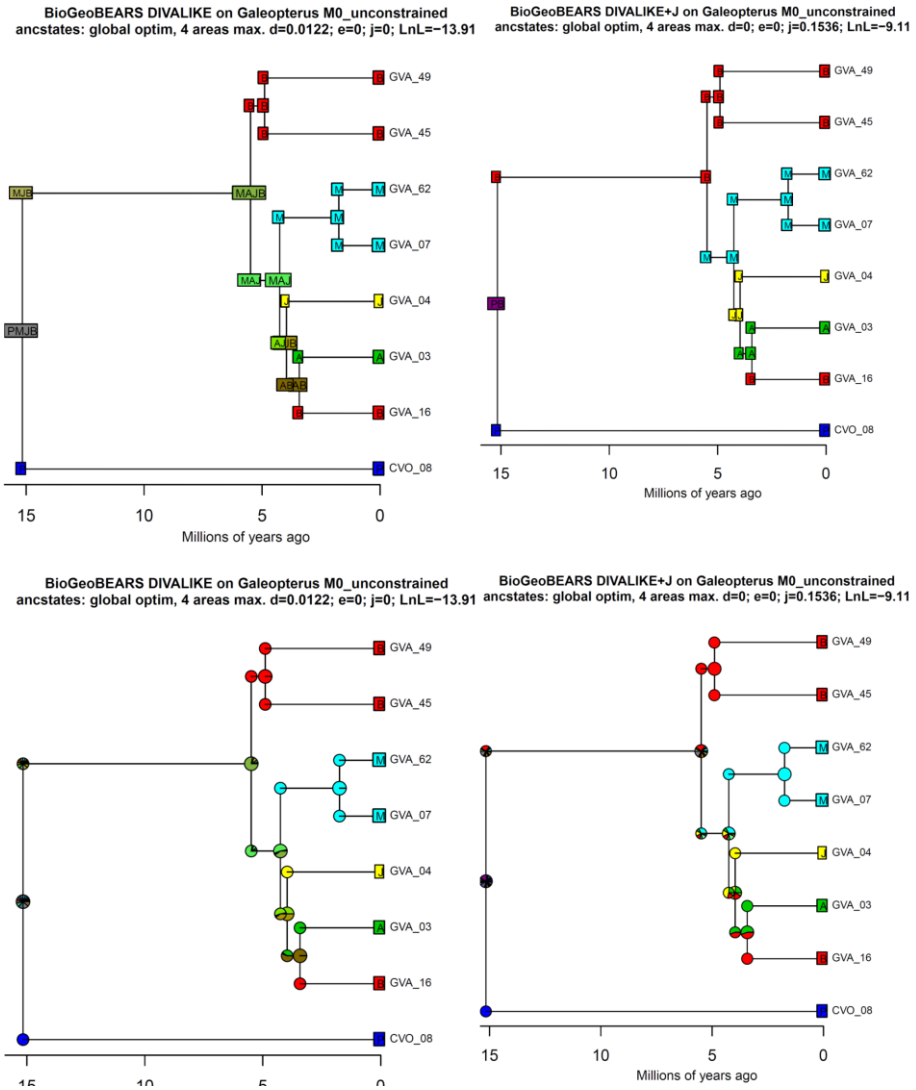


Figure A4.1 Continued.

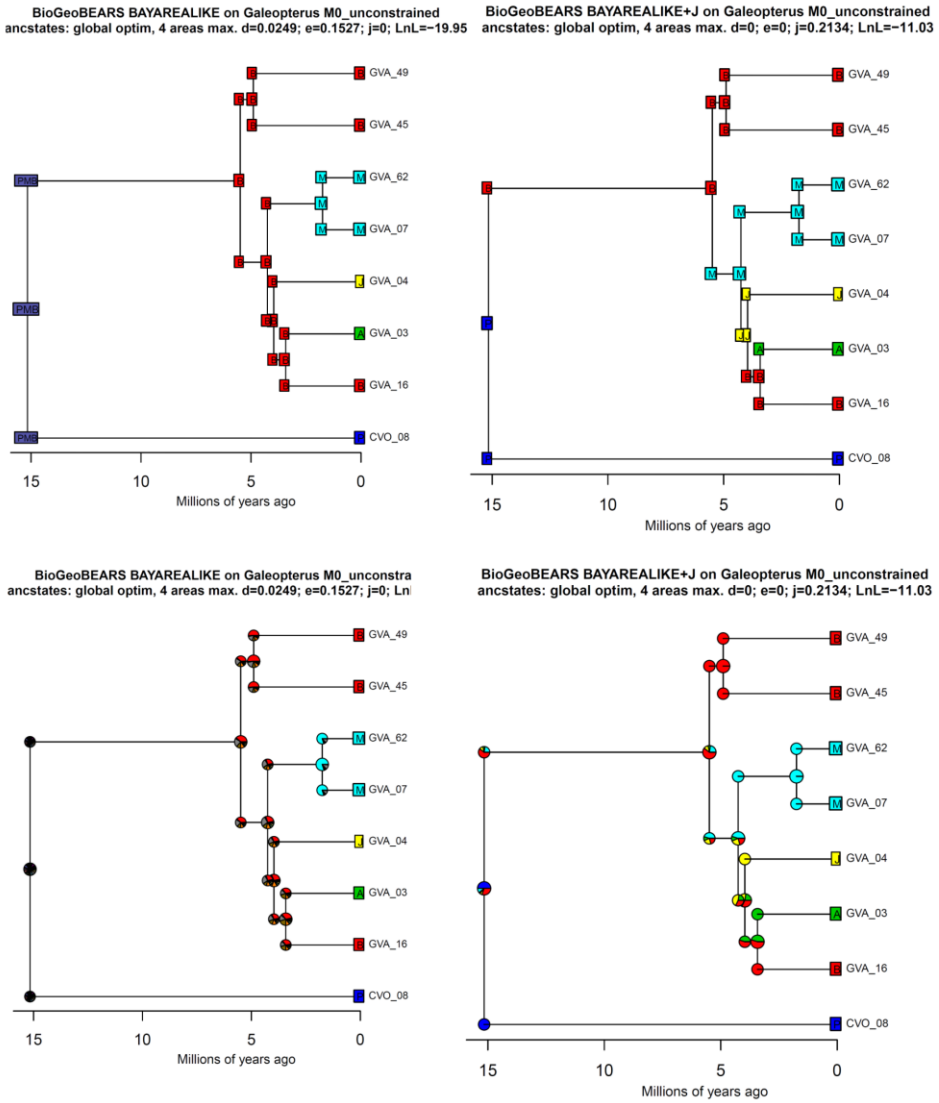


Figure A4.1 Continued.

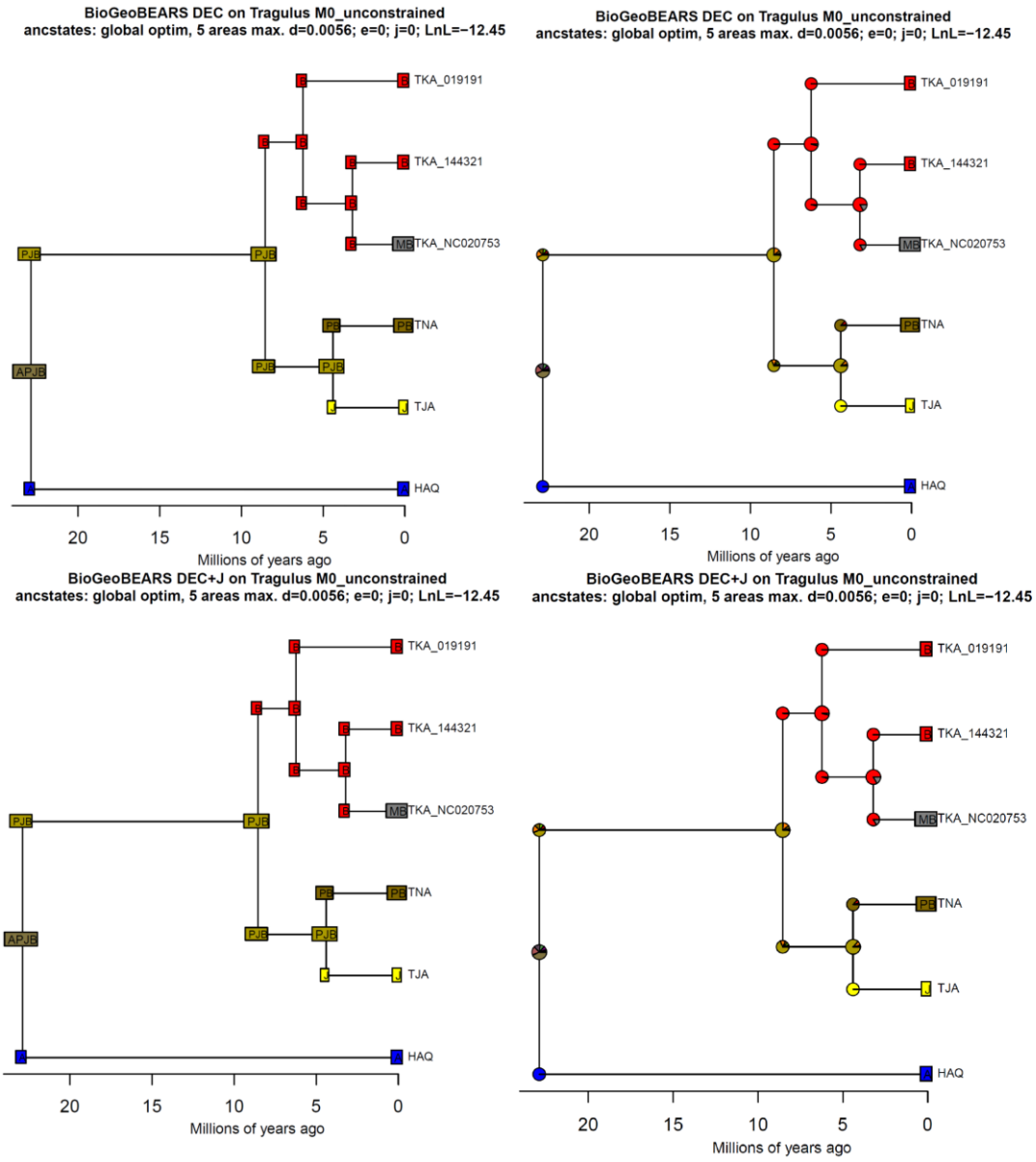


Figure A4.2 Ancenstral Area Probability calculations for the lesser, greater, and Javan mouse deer species with models DEC, DEC+j, DIVA, DIVA+j, BayArea, and BayArea+J.

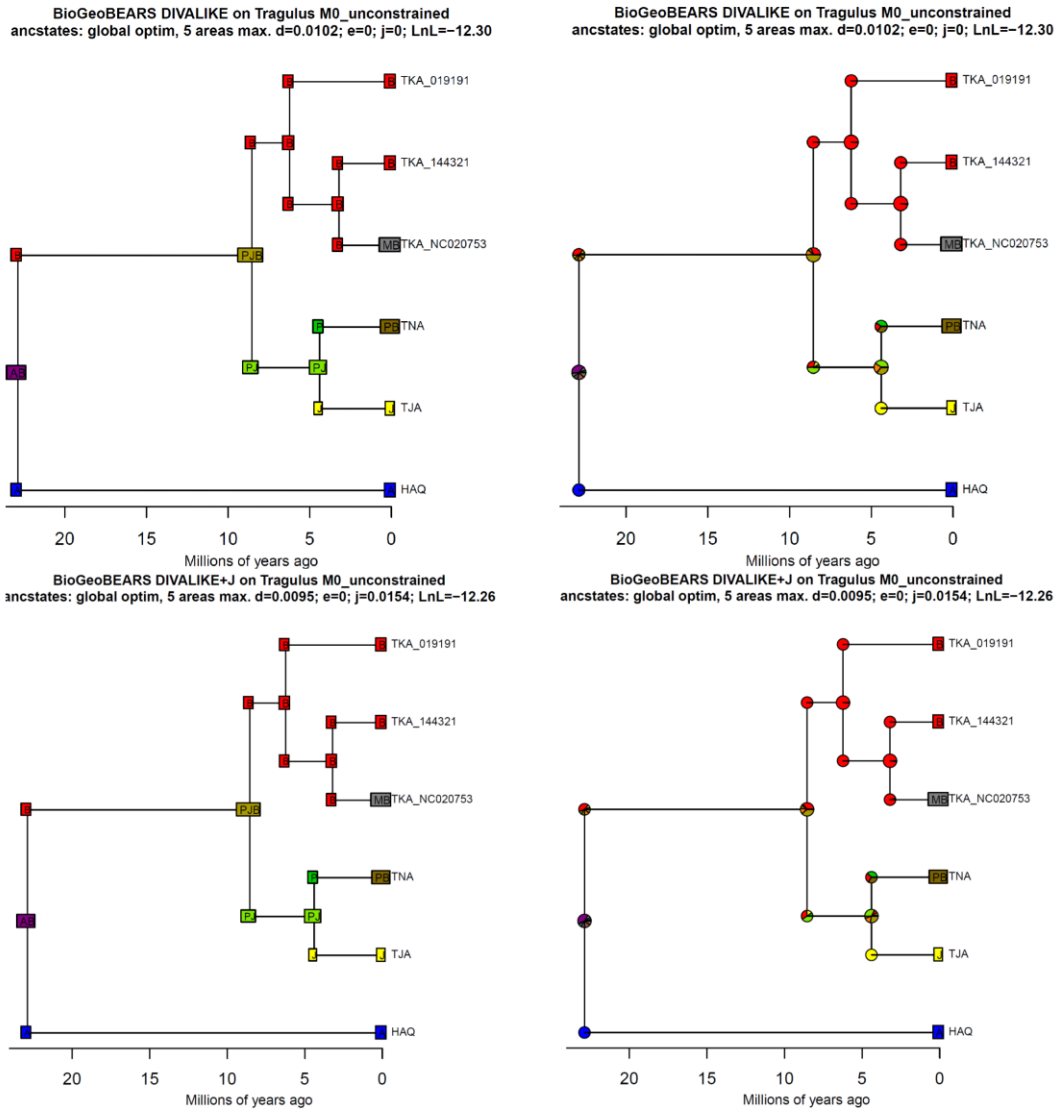


Figure A4.2 Continued.

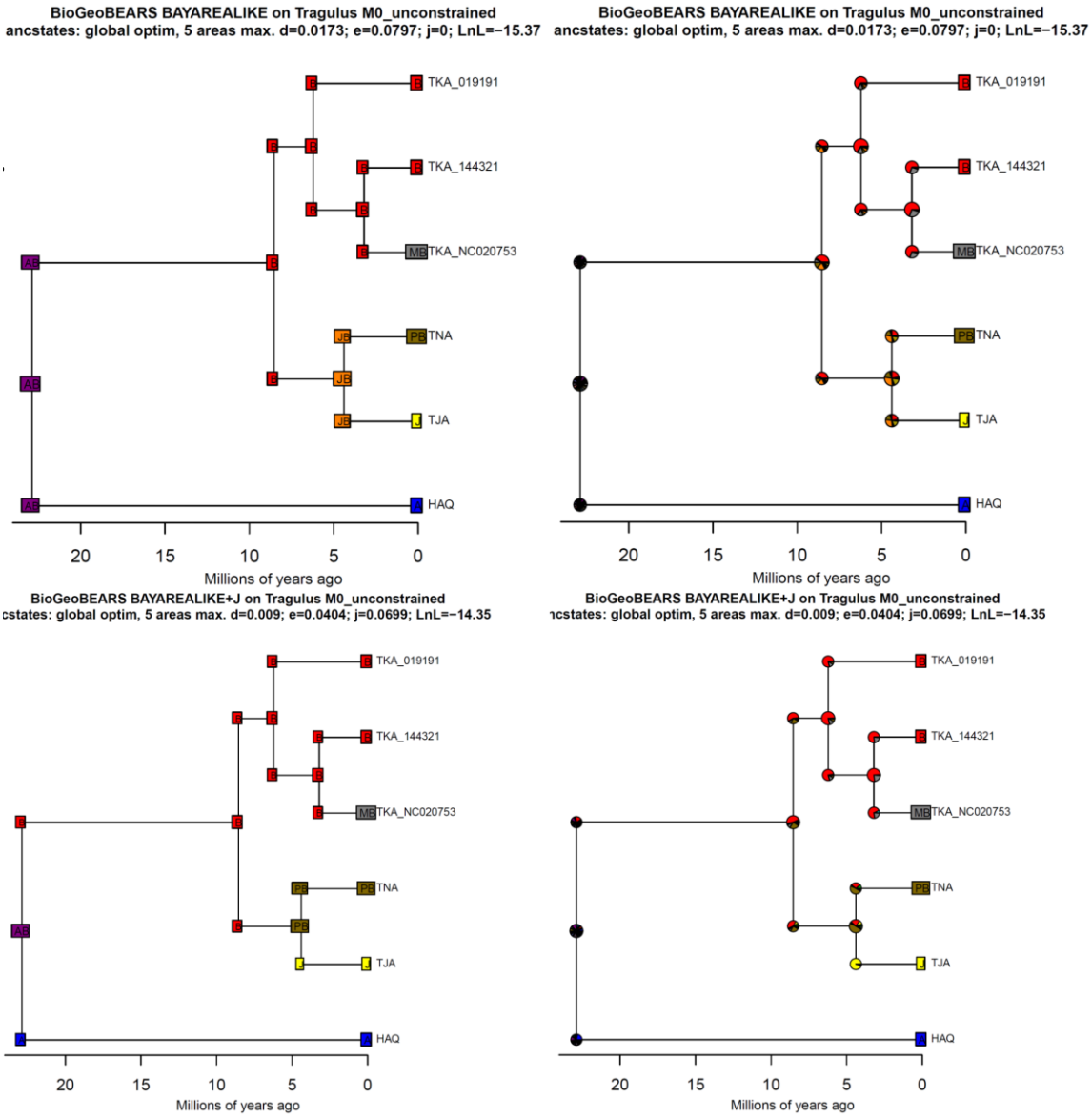


Figure A4.2 Continued.

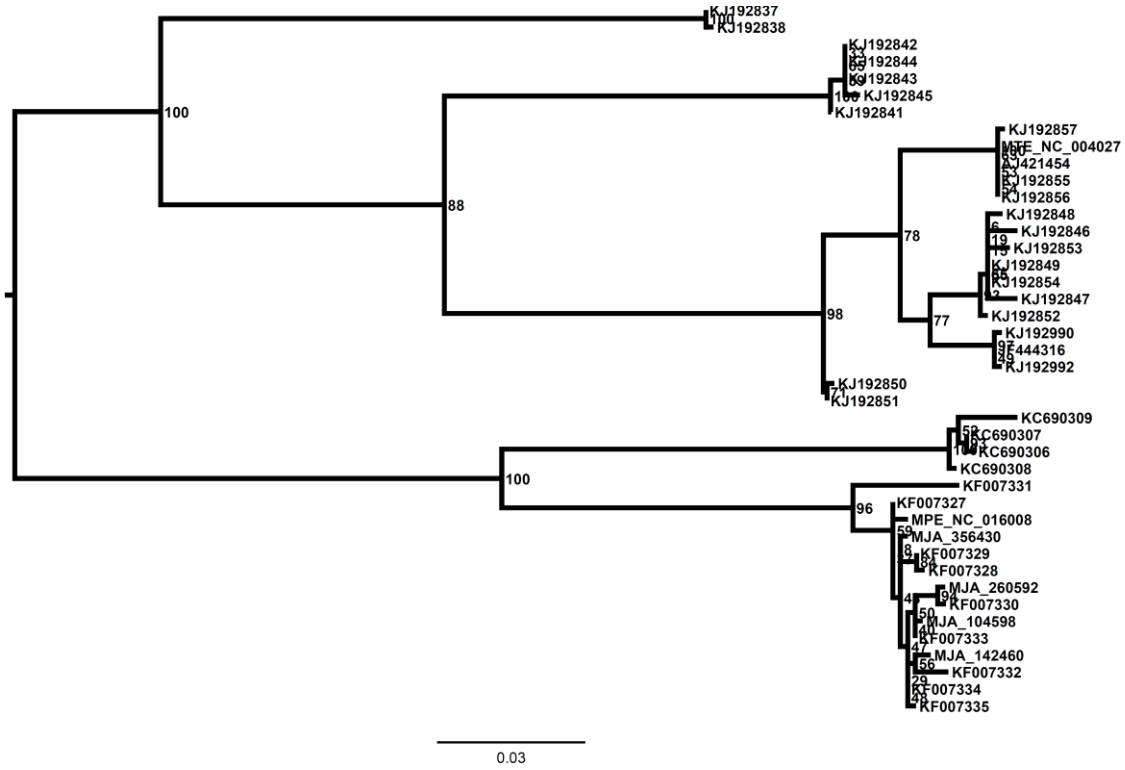


Figure A4.3. African pangolins and Sundaic pangolin COI ML phylogeny with sequences from NCBI. No COI sequence was recovered from E. Bornean MJA317198 for comparison.

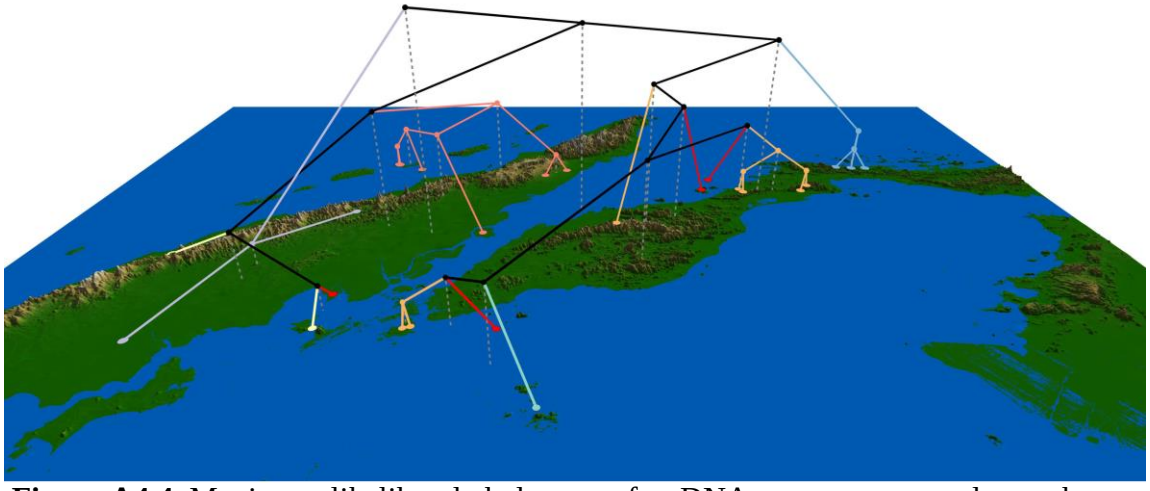


Figure A4.4. Maximum likelihood phylogeny of mtDNA genomes mapped onto the geography of Peninsular Malaysia and Sumatra (viewed from the east) showing microstructure and independent origins of dwarf populations. Dwarf individuals from dwarf populations are shown in red.

APPENDIX B

SUPPLEMENTAL TABLES

Table B2.1. Quantification of initial colugo DNA extracts

ID	ng/µL	260/280	Midpoint of DNA Smear
1	65.21	1.76	400
2	16.17	1.86	175
3	205.41	1.80	650
4	9.46	1.92	200
5	64.14	1.84	250
6	9.85	1.67	150
7	126.61	1.58	125
8	225.35	1.84	350
9	15.35	1.93	200
10	24.26	1.67	300
11	236.65	1.64	350
12	234.29	1.88	250
13	16.71	1.41	75

Table B2.2. Categorical classification of reads that did not align to the reference sequence.

I.D.	GVA mtDNA	Potential Numts	Other mtDNA	GVA Nuclear	Human Nuclear	Potential GVA Nuclear	Other Nuclear	Bacterial
1	0.06	2.83	2.75	0.17	0.00	0.05	0.00	0.00
2	6.46	0.00	3.47	0.00	0.59	0.66	1.20	0.54
3	3.85	0.96	3.10	0.00	0.00	0.00	0.03	0.00
4	2.04	0.20	1.14	0.00	0.00	0.00	0.00	0.00
5	4.95	0.93	1.00	0.00	0.00	0.00	0.00	0.00
6	20.11	0.00	2.80	0.01	0.00	3.68	0.16	0.00
7	2.78	0.00	1.95	0.00	0.26	2.49	1.05	2.24
8	5.19	0.82	0.46	0.00	5.80	0.00	0.09	0.30
9	0.00	0.00	16.21	0.00	0.00	0.02	10.40	0.41
10	4.50	0.70	2.77	0.01	2.69	0.00	0.15	0.00
11	5.34	0.01	0.96	0.00	0.00	0.00	0.04	0.02
12	3.05	0.78	1.73	0.02	0.00	2.37	0.00	0.28
13	0.09	0.00	0.00	0.00	4.23	0.58	0.00	0.34
Average:	4.49	0.56	2.95	0.27	0.78	0.76	1.01	0.32

Table B2.3. Average sequence depth per site. The first column lists the number of sites that meet the requirement of depth 5, the second lists the total number of nucleotides that meet the depth requirement, the third lists the depth/site.

Specimen	Number of Sites	Total Nucleotide Depth	Depth/site
1	15069	5826393	386.65
2	8040	2083649	259.16
3	15182	3736938	246.14
4	13058	2905341	222.50
5	4825	165591	34.32
6	10230	31392326	3068.65
7	3851	9103187	2363.85
8	3929	9493576	2416.28
9	2256	22605458	10020.15
10	15864	10165381	640.78
11	12035	4938501	410.34
12	14967	16563249	1106.65
13	2739	517622	188.98
ALL Individuals	122,045	119,497,212	979.12

Table B2.4. Assessment of chemical damage. *A*) The total number of transitions and transversions are shown for each individual, where "TC" indicates a transition from thymine in the reference sequence to cytosine in the specimen consensus sequence. *B*) The number of complementary base transitions CT + GA and TC + AG. A chi-square test was used to test for significance. *C*) The percentage of sites with CT and GA transitions in each consensus sequence at depth 5.

A.

Number of Directional Substitutions													
I.D.	Reference:	TC	TA	TG	CT	CA	CG	AT	AC	AG	GT	GC	GA
1	AF460846	51	2	0	57	0	0	1	2	22	3	1	16
2	AJ428849	157	6	3	189	10	0	11	10	85	3	3	75
3	AJ428849	221	5	4	244	9	2	9	7	132	2	2	131
4	AJ428849	165	3	1	209	10	1	9	6	101	2	1	99
5	AF460846	106	13	1	71	9	0	0	7	54	1	3	38
6	AF460846	280	13	3	190	14	1	11	9	135	1	4	108
7	AJ428849	74	5	5	91	6	2	6	4	42	0	1	37
8	AF460846	126	5	5	127	16	4	12	10	61	4	3	47
9	AJ428849	32	16	2	51	17	2	19	6	26	1	3	26
10	AJ428849	216	5	1	273	10	1	8	7	137	2	1	134
11	AJ428849	161	3	1	188	8	1	7	5	93	2	1	101
12	AJ428849	202	6	1	248	13	0	10	7	135	2	1	124
13	AF460846	68	2	1	47	4	1	1	6	39	0	2	24
	Total:	1859	84	28	1985	126	15	104	86	1062	23	26	960

Table B2.4 Continued.

B.

ID	CT + GA	TC + AG	P-value
1	73	73	1.000
2	264	242	0.328
3	375	353	0.415
4	308	266	0.080
5	109	160	0.002
6	298	415	<0.001
7	128	116	0.442
8	174	187	0.494
9	77	58	0.102
10	407	353	0.050
11	289	254	0.133
12	372	337	0.189
13	71	107	0.007
Total:	2945	2921	0.754

Table B2.4 Continued.

C.

ID	Total # of Sites	Total CT and GA Transitions	Percentage of Sites Affected by CT and GA
1	15099	73	0.48
2	7956	264	3.32
3	15141	375	2.48
4	13044	308	2.36
5	4807	109	2.27
6	10098	298	2.95
7	3809	128	3.36
8	4687	174	3.71
9	2292	77	3.36
10	15807	407	2.57
11	11819	289	2.45
12	14918	372	2.49
13	3167	71	2.24
Total	122644	2945	2.40
		Average:	2.62

Table B2.5. Open reading frame analysis.

ID	NADH1	NADH2	COI	COII	ATPase8	ATPase6	COIII	NADH3	NADH4L	NADH4	NADH5	NADH6	CYTB
1	2	1	1	3	2	2	1	1	2	4	2	1	4
2	4	3	4	3	4	5	5	3	4	5	5	5	3
3	1	2	1	2	1*	1	1	1	1	4	5	4	2
4	1	3	3	3	4	3	2	2	4	5	5	5	2
5	5	4	4	5	n/a	5	5	5	n/a	n/a	5	n/a	4
6	2*	2	2	3	4	3	2	3	4	5	5	5	5
7	4	n/a	3	5	n/a	5	n/a	3	n/a	5	5	n/a	5
8	5	5	n/a	3	4	3	4	n/a	n/a	3*	5	n/a	5
9	n/a	4	5	n/a	n/a	n/a	5	n/a	n/a	n/a	5	5	5
10	1	1	1	4	2	1	1	1	1	2	3	2	2
11	2	2	2	3	4	3	1	2	3	4	5	5	4
12	1	1	1	1	1	1	1	1	1	4	5	4	1
13	n/a	5	5*	n/a	n/a	5	5	n/a	n/a	5	5	5	5

LEGEND:
1 = Complete CDS with intact ORF (i.e. premature stop codon)
2 = Incomplete (i.e. gaps in) CDS, start and stop codons present, and intact partial ORF
3 = Incomplete CDS with the correct stop codon, no start codon present, and intact partial ORF
4 = Incomplete CDS with the correct start Codon, no stop codon present, and intact partial ORF
5 = Incomplete CDS with no start or stop codon present, and intact partial ORF
* = Premature stop codon in CDS
n/a = No sequence present

Fig. B2.6. B) SNP statistics for each individual.

Specimen	# of SNPs	# of AA altering SNPS	# of sites analyzed	% of sites w/minor allele ≥20%	% of SNP sites that are AA altering	% of SNPs that are AA altering
1	46	8	15099	0.305	0.053	17.39
2	44	7	7956	0.553	0.088	15.91
3	225	37	15141	1.486	0.244	16.44
4	25	12	13044	0.192	0.092	48.00
5	11	0	4807	0.229	0.000	0.00
6	58	12	10098	0.574	0.119	20.69
7	12	3	3809	0.315	0.079	25.00
8	5	0	4687	0.107	0.000	0.00
9	9	0	2292	0.393	0.000	0.00
10	40	20	15807	0.253	0.127	50.00
11	28	9	11819	0.237	0.076	32.14
12	43	11	14918	0.288	0.074	25.58
13	16	4	3167	0.505	0.126	25.00
Average:	43.23	9.46	9434.15	0.42	0.08	21.24

Table B2.7. Mitochondrial sequence divergence between Sunda colugos

	AJ428849	AF460846	1	2	3	4	5	6	7	8	9	10	11	12	13
AJ428849															
AF460846	0.124														
1	0.127	0.011													
2	0.074	0.067	0.065												
3	0.054	0.119	0.121	0.057											
4	0.050	0.124	0.126	0.052	0.018										
5	0.096	0.069	0.071	0.083	0.092	0.095									
6	0.113	0.086	0.088	0.093	0.106	0.107	0.043								
7	0.080	0.095	0.097	0.067	0.067	0.070	0.041	0.078							
8	0.137	0.100	0.099	0.121	0.130	0.142	0.034	0.063	0.059						
9	0.101	0.132	0.128	0.109	0.107	0.103	0.124	0.131	0.066	0.206					
10	0.051	0.125	0.128	0.053	0.019	0.012	0.094	0.114	0.069	0.139	0.105				
11	0.051	0.121	0.122	0.046	0.018	0.014	0.092	0.107	0.070	0.136	0.100	0.014			
12	0.052	0.127	0.129	0.046	0.021	0.013	0.098	0.112	0.069	0.141	0.103	0.013	0.014		
13	0.112	0.074	0.082	0.069	0.106	0.111	0.033	0.076	0.043	0.030	0.106	0.113	0.108	0.119	
GVA_4	0.125	0.003	0.011	0.067	0.121	0.125	0.070	0.087	0.097	0.101	0.130	0.127	0.122	0.129	0.076

The number of base substitutions per site between individual sequences are shown. Analyses were conducted using the Maximum Composite Likelihood model [1]. The analysis involved 16 nucleotide sequences. All positions with less than 50% site coverage were eliminated. That is, fewer than 50% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 14008 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [2]

1. Tamura K., Nei M., and Kumar S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (USA) 101:11030-11035.

2. Tamura K., Peterson D., Peterson N., Stecher G., Nei M., and Kumar S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution (In Press)

Table S2.8. Pairwise sequence divergence from the probe sequence

Approximate Genetic Divergence (ML distance) from Probe (GVA4)							
(5,459bp)70%*			(14,008bp)50%*			(16,051bp)10%*	
0.2%	AF		0.3%	AF		0.3%	AF
1.0%	1		1.1%	1		1.1%	1
6.4%	2		6.7%	2		6.6%	2
6.6%	5		7.0%	5		7.0%	5
7.7%	6		7.6%	13		7.4%	13
7.9%	13		8.7%	6		8.8%	6
9.1%	7		9.7%	7		9.8%	7
9.5%	3		10.1%	8		10.4%	8
9.8%	AJ		12.1%	3		12.3%	11
9.9%	11		12.2%	11		12.4%	3
10.0%	4		12.5%	4		12.6%	4
10.0%	10		12.5%	AJ		12.9%	AJ
10.0%	12		12.7%	10		13.0%	10
10.2%	8		12.9%	12		13.0%	9
11.8%	9		13.0%	9		13.1%	12
	Legend: Phylogenetic groupings						
	W. Java						
	E. Java						
	Borneo-1/Malaya						
	Pen.Malaysia/Sumatra						
	Natuna Islands						
	Borneo-2						
*The length of the alignment matrix based on sites for which ≥X% of the individuals possess a base at that site.							

Table S2.9. Primer pairs used to amplify Sunda colugo mtDNA genome probe fragments.

			Reference Seq: AJ428849		
	Start	End	Forward Primer Sequence (start)	Reverse Primer Sequence (end)	Size (bp)
1	25	1084	GCAAGGTACTGAAAATACCAAGATG	TGAAATCTTCCGGGTGTAGG	1060
2	847	1958	CAAAGGAGGATTTAGCAGTAAATTAAG	TGCTAGAGGTGATGTTTTTGG	1112
3	1621	2681	GCCACCAATTAAGATAGCGTTC	CTAACAAGCCCTGCTCTTGG	1061
4	2426	3809	CTCGATGTTGGATCAGGACA	TTCTCAGGAGTGGGTTCGAT	1382
5	3573	4328	CGAGCTTCATACCCACGATT	GGCTAGTTTTTGTCATGTCAGG	756
6	4062	5340	AACCCACGATCAACAGAAGC	AGGGTGAGGTGGCTGAGTAA	1279
7	5171	6473	CTACTTCTCCCGCCTCCAAG	TGTGCTACTACGTAATATGTGTCGTG	1303
8	6319	7221	GCTACACTGCACGGAGGAA	TGGTTTCTACTATTTGGGCATTT	903
9	6975	8348	AAAAACATTACATGACTTCGTCAGA	GGTGTGCCTTGGGGTAGAAG	1374
10	7777	8951	CCACAATGAAATGCCACAAC	TGGAGCTAGGCTTGAGTGGT	1175
11	8559	9664	CACCGTAGCCCTAATCCAAG	ACGTGATGGCCACTAGGAAA	1106
12	8864	9867	ACGATACGGAATAATTCTCTTCA	AATGGGTCGAAACCAGTTGT	1004
13	9628	10837	CCCTTCTCCATAAAATTTTTCC	TTTTGGTAGTCAGAGGTGAAGTC	1210
14	10550	11697	GAAGCAACACTAATCCCAACC	TTGAAAGTAAGAAAGCCATATTTTT	1148
15	11334	12602	CAGCATTCTCCTGATCAAACA	AGTGTGGTGAGGGCACCTA	1269
16	12431	13869	TACACCCGTGACTTCCCTCT	TACTGCCATGGCTATTGAGG	1439
17	13660	14806	GTAGAATCCCCATGAAAATAACC	GGGATTTTGTCTGAGTTTGATG	1147
18	14660	15919	AGACAAAGCCACCCTCACAC	GCATGGCCCTGAAGTAAGAA	1260
19	15349	16734	CTCCCCAGGACAATCAAGG	GCTTCAGGCCAAAATTCAAA	1386

Table B3.1. Boreoeutherian mammals used in phylogenetic comparisons

Common Name	Species Name	Sequence Origin
Human	*Homo sapiens*	Ensembl v.79
Chimpanzee	*Pan troglodytes*	Ensembl v.79
Gorilla	*Gorilla gorilla*	Ensembl v.79
Orangutan	*Pongo abelli*	Ensembl v.79
Macaque	*Macaca fascicularis*	Ensembl v.79
Marmoset	*Callithrix jacchus*	Ensembl v.79
Bushbaby	*Otolemur garnetti*	Ensembl v.79
Sunda colugo	*Galeopterus variegatus*	G_variegatus-3.0.2
Philippine colugo	*Cynocephalus volans*	14x ref. assembly-this study
Chinese treeshrew	*Tupaia belangeri chinensis*	(*48*)
Pen-tailed treeshrew	*Ptilocercus lowii*	5x ref. assembly-this study
Rabbit	*Oryctolagus cuniculus*	Ensembl v.79
13-lined ground squirrel	*Ictidomys tridecemlineatus*	Ensembl v.79
Guinea Pig	*Cavia porcellus*	Ensembl v.79
Rat	*Rattus rattus*	Ensembl v.79
Mouse	*Mus musculus*	Ensembl v.79
Cow	*Bos taurus*	Ensembl v.79
Cat	*Felis catus*	Ensembl v.79
Dog	*Canis familiaris*	Ensembl v.79
African elephant	*Loxodonta africana*	Ensembl v.79

Table B3.2. Chi-square calculation for phylogenetically informative indels (Waddell 2001).

	Observed	**Expected**	**chi-square**	
Primatomorpha	19	9	11.1	
Sundatheria	4	9	2.8	
Primates+Tupaia	4	9	2.8	
Sum(values)	27	27	16.7	
***P*-value**			0.000045	df = 1

Table B3.3. Size of Olfactory receptor (*OR*) and vomeronasal class 1 receptor (*V1R*) gene families found in colugo and 16 select mammalian genomes.

Node	*OR* gene families			*V1R* gene families		
	intact genes	**partial/ pseudo genes**	**Repertoire size**	**intact genes**	**Partial /pseudo genes**	**Repertoire size**
human [REF]	396	425	821	5	115	120
orangutan	296	525	821	5	114	119
rhesus	319	321	640	0	51	51
marmoset	366	258	624	8	42	50
galago	356	585	941	60	49	109
mouse lemur	361	619	980	82	109	191
colugo	518	1480	1998	46	107	153
Chinese treeshrew	969	1396	2365	23	55	78
mouse [REF]	1127	202	1329	187	121	308
rat [REF]	1194	438	1632	106	66	172
rabbit	751	278	1029	160	132	292
cat	679	330	1009	21	47	68
dog	811	246	1057	8	29	37
horse	1063	1511	2574	36	47	83
pig	1254	782	2036	8	28	36
cow	1055	926	1981	40	24	64
opossum	1157	300	1457	100	26	126

Table B3.4. Nuclear capture efficiency and comparison of individual versus multiplex capture results. Multiplex and Individual nuclear capture columns report the percentage of reads mapped to reference probe sequences after removal of single end and paired end duplicates and merging of overlapping read pairs. Indiv. / Multiplex column reports the relative fold-increase in efficiency when performing single individual capture compared to multiple individual capture.

Specimen	Multiplex Nuclear Capture	Individual Nuclear Capture	Indiv. / Multiplex
CVO_06	0.14	0.14	1.0
CVO_07	0.01	0.02	2.0
CVO_08	0.20	0.75	3.8
CVO_15	0.14	0.14	1.0
CVO_22	0.05	0.16	3.2
CVO_24	0.02	0.09	4.5
GVA_12	0.04	0.09	2.6
GVA_17	0.02	0.05	2.5
GVA_22	0.07	0.08	1.1
GVA_26	0.15	0.22	1.5
GVA_27	0.09	0.17	1.9
GVA_32	0.08	0.17	2.1
GVA_40	0.13	0.21	1.6
GVA_49	0.08	0.15	1.9
GVA_61	0.04	0.07	1.8
GVA_64	0.01	0.04	4.0
GVA_75	0.03	0.07	2.3
GVA_76	0.03	0.09	3.0
GVA_129	0.10	0.10	1.0
Average:	0.08	0.15	2.2

Table B3.5. Adapter blocking oligos partially derived from Maricic (2010).

Adapter Blocked	Blocking Oligo Sequence
#1 Blocks P5:	5' TGTAGATCTCGGTGGTCGCCGTATCATT-P 3'
#2 Blocks Rd1 SP:	5' AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-P 3'
#3 Blocks Rd2 SP:	5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-P 3'
#4 Blocks P7:	5' CAAGCAGAAGACGGCATACGAGAT-P 3'
#5 Blocks P5 (Complement)	5' AATGATACGGCGACCACCGAGATCTACAC-P 3'
#6 Blocks Rd1 SP (Complement)	5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT-P 3'
#7 Blocks Rd2 SP (Complement)	5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-P 3'
#8 Blocks P7 (Complement)	5' ATCTCGTATGCCGTCTTCTGCTTG-P 3'

Table B3.6. mtDNA reference assembly statistics from off-target nuclear capture reads.

Individual	Reads Mapped	Total Reads	Average Depth	% of Reference Bases Covered
CVO_06	1683	3315220	11.81	95.15
CVO_07	5683	9655652	36.07	96.64
CVO_08	3758	6397053	30.13	95.93
CVO_10	4643	10610200	23.89	95.81
CVO_13	146	1740932	0.75	43.55
CVO_15	4587	8669257	24.08	96.85
CVO_21	2096	4369074	8.24	95.44
CVO_22	12524	10240636	64.23	99.40
CVO_24	647	9673217	4.67	89.58
GVA_07	5292	6220128	39.63	98.85
GVA_09	1059	29808549	4.79	66.96
GVA_12	12411	14150951	81.37	99.95
GVA_16	6509	12224335	23.31	99.73
GVA_17	8125	6930959	54.78	99.97
GVA_18	6171	14485997	30.26	99.85
GVA_22	8263	3584858	21.87	99.70
GVA_27	364	1493420	1.41	73.37
GVA_28	650	7673989	2.12	82.41
GVA_32	2863	5050725	13.53	98.73
GVA_35s	2669	11384773	11.71	99.18
GVA_39s	13891	6333908	63.6	99.98
GVA_40	7273	8461912	41.68	99.55
GVA_45	16923	13081551	94.3	100.00
GVA_49	11396	5959804	91.28	100.00
GVA_52	103	6106934	0.52	37.50
GVA_55	4984	15265857	26.69	99.75
GVA_61	1392	1551730	9.12	97.92
GVA_62	4794	16809740	25.36	96.46
GVA_63	273	7911650	0.92	48.95
GVA_64	17615	11341089	79.11	99.99
GVA_69	2137	17168303	6.77	96.57
GVA_71	142	3077004	0.35	25.05
GVA_75	337	2478753	1.11	60.87
GVA_76	2314	9911491	7.18	99.05
GVA_78	36	5498695	0.12	10.72
GVA_106	64	4035807	0.25	18.01
GVA_115	1002	5719343	3.58	92.03
GVA_121	20	3982512	0.06	6.06
GVA_129	20842	12079124	162.86	99.99
GVA_133	570	5284258	2.36	76.68
GVA_134	3530	13585155	22.13	99.92
Mean:	4873	8617672	28.89	84.77
Median:	2863	7673989	21.87	96.85

Table B3.7. External fossil calibrations used to calculate the divergence time between colugo genera. In some cases original calibrations have been updated to correspond to revisions in the Stratigraphic Code.

Pair of Taxa	Lower Bound (mya)	Upper Bound (mya)	Reference
Dog-Cat (*Canis-Felis*)	38	66	Emerling 2015
Rabbit-Pika (*Oryctolagus-Ochotona*)	53.7	61.6	Minimum based on Ypresian tarsals of crown lagomorph (Rose 2008) with age of 53.7 Ma (Kapur 2015); maximum following (Emerling 2015)
Macaque-Marmoset (*Macaca-Callithrix*)	28.1	56	Emerling 2015
Macaque-Human (*Macaca-Homo*)	20.55	38	Emerling 2015
Primates	56	66	Benton et al. 2015
Primatomorpha	65.2	83.8	Minimum based on *Purgatorius* (oldest crown fossil, (Chester et al. 2015); maximum based on stratigraphic bounding as in Meredith et al. (2011) with two stages
Euarchontoglires-Laurasiatheria	65.2	131.5	Minimum based on *Purgatorius* (Chester et al. 2015); maximum based on Benton et al., (Benton et al. 2009)

Table B3.8. Maximum mtDNA genetic divergence between seven Sunda colugo groups and five Philippine colugo groups, based on the composite likelihood + gamma genetic distance. Abbreviations: Vietnam=VNM, Thailand=THA, Peninsular Malaysia=PM, Sumatra=SUM, Borneo=BOR, Pulau Bunguran=BUN, Pulau Basilan=BAS, Mindanao=MIN, Bohol=BOL, Leyte=LTE, Samar=SAM, Pulau Dinagat=DIN.

	PM THA SUM	W.BOR	JAVA	LTE SAM	BOL	E.BOR	NE.BOR	E.MIN	W.MIN	LAOS	VNM
THA/PM/SUM											
W.BOR	0.058										
JAVA	0.153	0.149									
LTE/SAM	0.216	0.210	0.220								
BOL	0.216	0.210	0.222	0.050							
E.BOR	0.150	0.146	0.121	0.216	0.216						
NE.BOR	0.144	0.141	0.121	0.211	0.212	0.070					
E.MIN	0.212	0.211	0.216	0.048	0.052	0.215	0.209				
W.MIN	0.211	0.209	0.216	0.050	0.053	0.214	0.207	0.035			
LAOS	0.080	0.076	0.151	0.210	0.209	0.145	0.141	0.210	0.207		
VNM	0.070	0.069	0.144	0.196	0.197	0.136	0.133	0.197	0.196	0.064	
DIN	0.209	0.201	0.215	0.032	0.041	0.212	0.208	0.041	0.040	0.202	0.190
					Mean	**Stdev**	**Min**	**Max**	**Median**		
Average Sundaic Between Group Divergence					0.117238	0.035714	0.058	0.153	0.136		
Average Philippine Between Group Divergence					0.0442	0.007391	0.032	0.053	0.0445		

Table B3.9. X-Chromosome sequence divergence between eight Sunda colugo groups and three Philippine colugo groups, based on maximum composite likelihood + gamma genetic distance. Includes all individuals with >40% probe base coverage. Abbreviations: Vietnam=VNM, Thailand=THA, Peninsular Malaysia=PM, Sumatra=SUM, Borneo=BOR, Pulau Bunguran=BUN, Pulau Basilan=BAS, Mindanao=MIN, Bohol=BOL, Leyte=LTE, Samar=SAM, Pulau Dinagat=DIN.

	E.BOR	THA/PM/ SUM	VNM	NW.BOR	S.SUM	LTE/SAM	NE.BOR	E. MIN	JAVA	DIN	SE.BOR
E.BOR											
THA/PM SUM	0.0070										
VNM	0.0057	0.0056									
NW.BOR	0.0051	0.0052	0.0039								
S.SUM	0.0051	0.0039	0.0043	0.0034							
LTE/SAM	0.0276	0.0291	0.0288	0.0289	0.0274						
NE.BOR	0.0041	0.0073	0.0059	0.0059	0.0057	0.0285					
E.MIN	0.0267	0.0285	0.0282	0.0280	0.0270	0.0014	0.0276				
JAVA	0.0062	0.0062	0.0052	0.0051	0.0047	0.0299	0.0069	0.0291			
DIN	0.0285	0.0301	0.0299	0.0302	0.0281	0.0010	0.0294	0.0012	0.0312		
SE.BOR	0.0038	0.0096	0.0080	0.0073	0.0071	0.0311	0.0053	0.0296	0.0089	0.0315	
LAOS	0.0060	0.0058	0.0027	0.0047	0.0047	0.0294	0.0066	0.0286	0.0056	0.0302	0.0086
					Mean	**Stdev**	**Min**	**Max**	**Median**		
Average Sundaic Between Group Divergence					0.0058	0.0016	0.0027	0.0096	0.0057		
Average Philippine Between Group Divergence					0.0012	0.0002	0.0010	0.0014	0.0012		

Table B3.10. Pairwise sequence divergence between colugos from Peninsular Mainland, Sumatra and Thailand, and their satellite islands (blue).

	THA	P.Bakong	P.Tanahbala	P.Aur	P.Siantan	PM	P.Langkawi	P.Bintang	KoAdang
THAILAND									
P.Bakong	0.0109								
P.Tanahbala	0.0110	0.0102							
P.Aur	0.0106	0.0109	0.0104						
P.Siantan	0.0097	0.0101	0.0096	0.0078					
PEN.MALAY.	0.0113	0.0109	0.0112	0.0107	0.0103				
P.Langkawi	0.0104	0.0106	0.0112	0.0101	0.0094	0.0102			
P.Bintang	0.0099	0.0102	0.0102	0.0032	0.0070	0.0097	0.0094		
KoAdang	0.0084	0.0103	0.0097	0.0114	0.0114	0.0099	0.0112	0.0108	
P.Pini	0.0104	0.0097	0.0030	0.0103	0.0094	0.0100	0.0104	0.0098	0.0105
SUMATRA	0.0111	0.0102	0.0101	0.0114	0.0106	0.0111	0.0112	0.0106	0.0109
P.Penuba	0.0110	0.0025	0.0103	0.0112	0.0103	0.0117	0.0107	0.0104	0.0105
P.Rupat	0.0111	0.0097	0.0111	0.0111	0.0108	0.0121	0.0107	0.0115	0.0118
					Mean	**Stdev**	**Min**	**Max**	**Median**
THA/PM/SUM to island sequence divergence					0.0105	0.0007	0.0084	0.0121	0.0106

Table B3.10 Continued.

	P.Pini	SUM	P.Penuba
THAILAND			
P.Bakong			
P.Tanahbala			
P.Aur			
P.Siantan			
PEN.MALAY.			
P.Langkawi			
P.Bintang			
KoAdang			
P.Pini			
SUMATRA	0.0095		
P.Penuba	0.0094	0.0103	
P.Rupat	0.0095	0.0105	0.0100

Table B3.11 Within-group mtDNA maximum composite likelihood + gamma genetic distance matrix for nine Sunda colugo groups and five Philippine colugo groups. Abbreviations: Vietnam=VNM, Thailand=THA, Peninsular Malaysia=PM, Sumatra=SUM, Borneo=BOR, Pulau Bunguran=BUN, Pulau Basilan=BAS, Mindanao=MIN, Bohol=BOL, Leyte=LTE, Samar=SAM, Pulau Dinagat=DIN.

THA/PM/SUM	0.010
SW. BOR	0.007
JAVA	0.007
NW.BOR	0.003
SAM/LTE	0.003
BOL	0.000
E.BOR	0.011
NE.BOR	0.006
E.MIN	0.000
Zamboanga	0.002
S.SUM	0.006
LAOS	n/c
VNM	n/c
DIN	n/c

Table B3.12. mtDNA genetic distance between nine Sundaic colugo groups and five Philippine colugo groups maximum composite likelihood + gamma distance matrix. Abbreviations: Vietnam=VNM, Thailand=THA, Peninsular Malaysia=PM, Sumatra=SUM, Borneo=BOR, Pulau Bunguran=BUN, Pulau Basilan=BAS, Mindanao=MIN, Bohol=BOL, Leyte=LTE, Samar=SAM, Pulau Dinagat=DIN.

	PM THA SUM	SW. BOR	JAVA	NW. BOR	LTE SAM	BOL	E. BOR	NE. BOR	E. MIN	W. MIN	S. SUM	LAOS	VNM
PM THA SUM													
SW.BOR	0.059												
JAVA	0.154	0.149											
NW.BOR	0.058	0.037	0.149										
LTE/SAM	0.217	0.208	0.22	0.210									
BOL	0.217	0.209	0.222	0.210	0.050								
E.BOR	0.151	0.145	0.121	0.146	0.216	0.216							
NE.BOR	0.146	0.140	0.121	0.141	0.211	0.212	0.07						
E.MIN	0.213	0.208	0.216	0.212	0.048	0.052	0.215	0.209					
W.MIN	0.213	0.206	0.216	0.210	0.050	0.053	0.214	0.207	0.035				
S.SUM	0.034	0.057	0.146	0.057	0.204	0.205	0.138	0.136	0.200	0.199			
LAOS	0.081	0.075	0.151	0.076	0.210	0.209	0.145	0.141	0.210	0.207	0.074		
VNM	0.070	0.067	0.144	0.070	0.196	0.197	0.136	0.133	0.197	0.196	0.070	0.064	
DIN	0.210	0.202	0.215	0.201	0.032	0.041	0.212	0.208	0.041	0.040	0.193	0.202	0.190
							mean	stdev	min	max	median		
Average Sundaic Between Group Divergence							0.107	0.040	0.034	0.154	0.127		
Average Philippine Between Group Divergence							0.044	0.007	0.032	0.053	0.045		

Table B3.13. Morphometric principal component loadings for males and females with no dwarfs.

	PC1	PC2	PC3
GSL	0.18932	-0.09443	0.17312
CBL	0.18564	-0.10606	0.19291
ONL	0.20692	-0.14089	0.21679
zyg	0.18509	-0.07422	0.22832
min.w.temps	0.57311	0.797471	-0.06760
P4.M3	0.17849	-0.15021	-0.11203
bbc	0.16389	-0.01663	0.28649
hbc	0.11709	-0.02596	0.06355
P4W	0.21061	-0.24898	-0.13606
M1W	0.24827	-0.16391	-0.21752
M2W	0.26188	-0.18617	-0.20261
M3W	0.22163	-0.16353	-0.10616
max.pal.w	0.22208	-0.09415	0.32784
diam.orbit.d.V.	0.15018	-0.01950	0.11652
palate.across.T2s	0.16431	-0.14817	0.09456
p4.m3	0.18205	-0.13435	-0.01946
mand.condyle.w	0.18609	-0.23571	0.30417
t1longL	0.16841	-0.12435	-0.51270
t2longL	0.18179	-0.15644	-0.35573

Table B3.14. Morphometric principal component loadings for males and females (dwarfs included) after normalizing by body size.

	PC1	PC2	PC3
GSL	-0.01663	0.017175	-0.00442
ONL	0.023339	-0.02825	-0.12767
zyg	-0.05148	0.062697	0.183351
min.w.temps	-0.96799	-0.15036	-0.07986
P4.M3	-0.0411	0.290691	0.060051
bbc	-0.11283	0.083246	0.14261
hbc	-0.10536	0.167393	0.146996
P4W	0.012156	0.372447	0.076299
M1W	-0.07394	0.385925	0.085834
M2W	-0.07768	0.382301	0.140197
M3W	-0.07258	0.381536	0.134362
max.pal.w	-0.04104	0.006042	0.125325
diam.orbit..d.V.	-0.10677	0.127659	0.206659
palate.across.T2s	0.002943	0.120638	-0.04047
p4.m3	-0.04006	0.221703	0.073767
mand.condyle.w	0.04722	0.122476	0.325222
t1longL	-0.01239	0.285843	-0.73167
t2longL	-0.00268	0.311116	-0.37502

Table B3.15. X-chromosome genetic variation principal component loadings.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
GVA_62	0.266233	0.547404	-0.22344	0.318666	-0.12787	0.199536	-0.22883	-0.17031	0.491531	-0.11319	-0.05209
GVA_55	0.106196	-0.07303	0.027003	-0.1837	-0.03574	0.083016	0.033093	-0.13947	-0.01564	-0.09455	0.90977
GVA_35s	-0.417930	-0.36815	-0.59937	0.276448	-0.35334	0.000792	0.174165	-0.02659	0.021265	0.119834	-0.0103
GVA_40	0.161232	-0.41209	0.569543	0.623993	0.013976	0.037057	-0.00685	-0.01491	0.043918	0.046969	-0.03364
GVA_49	-0.108230	-0.19643	-0.20001	-0.10953	0.7248	0.034303	-0.42293	-0.00116	0.130662	0.300926	-0.03684
GVA_45	-0.749030	0.418413	0.412216	-0.0457	0.004591	0.010241	0.0647	-0.02743	0.01838	0.055598	0.006742
GVA_07	0.234199	0.185082	-0.07773	0.00447	0.290208	-0.3249	0.710371	0.291528	0.131815	0.155633	-0.02253
GVA_63	0.189492	0.308727	-0.11744	0.165388	-0.03705	0.070891	-0.14557	0.061808	-0.81878	0.202542	-0.03823
GVA_16	0.109996	-0.05044	0.094384	-0.24809	-0.35616	-0.73628	-0.37578	0.054117	0.088315	0.068156	-0.07762
GVA_12	-0.06196	-0.09647	-0.08403	-0.03419	0.206989	-0.0607	-0.01122	0.091066	-0.16144	-0.88744	-0.16348
GVA_129	0.160713	-0.13246	0.088955	-0.38649	-0.06008	0.171362	0.252997	-0.70848	-0.04948	0.071862	-0.33086
GVA_28	0.109087	-0.13056	0.109916	-0.38127	-0.27032	0.514682	-0.04415	0.589822	0.119453	0.07366	-0.15091

Table B3.16. BPP species estimation (Yang 2015) results for the *Galeopterus* biparental dataset (53 partitions, with 9,628 sites after removal of sites with missing data, clean data=1).

Model	Nodes[a]	Posterior Probability	Posterior Probability	Posterior Probability	Posterior Probability	Posterior Probability
1	000000	0.00000	0.00000	0.00000	0.00000	0.00000
2	100000	0.00000	0.00000	0.00000	0.00000	0.00000
3	101000	0.00000	0.00000	0.00000	0.00000	0.00000
4	101010	0.00000	0.00000	0.00000	0.00000	0.00000
5	101011	0.00028	0.00000	0.00000	0.00000	0.00000
6	101100	0.00000	0.00000	0.00000	0.00000	0.00000
7	101110	0.00000	0.00004	0.00000	0.00000	0.00211
8	101111	0.00007	0.00001	0.00006	0.00000	0.00145
9	110000	0.00000	0.00000	0.00000	0.00575	0.04417
10	111000	0.00000	0.00000	0.00000	0.00989	0.00496
11	111010	0.00595	0.00199	0.00099	0.15714	0.06777
12	111011	0.42029	0.17070	0.16314	0.82608	0.64262
13	111100	0.00000	0.00000	0.00000	0.00000	0.00092
14	111110	0.01321	0.00720	0.00590	0.00006	0.02099
15	111111	0.56021	0.82007	0.82993	0.00109	0.21502
Parameters		Parameters estimated from data				
theta (α, β)		2, 408	2, 1000	2, 1000	2, 10	2, 10
tau (α, β)		2, 344	2, 10	2, 1000	2, 10	2, 1000

[a]The ancestral nodes (left to right) for *Galeopterus*: **1)** East Borneo to Northeast Borneo, **2)** East Borneo+NE Borneo, **3)** Node 4 to Node 5, **4)** Laos+Vietnam, **5)** Java to node 6, **6)** West Borneo to Pen.Malay./Sumatra.

Guide tree with PP for each node (i.e., support for distinct species) in the five analyses shown above (left to right):

((EB, NEB) #0.999650,((V, L) #0.573490, (J, (W, PS) #0.980845) #1.000000) #1.000000) #1.000000;
((EB, NEB) #0.999955,((V, L) #0.827310, (J, (W, PS) #0.990770) #1.000000) #1.000000) #1.000000;
((EB, NEB) #0.999945,((V, L) #0.835875, (J, (W, PS) #0.993115) #1.000000) #1.000000) #1.000000;
((EB, NEB) #1.000000,((V, L) #0.001145, (J, (W, PS) #0.827160) #0.984355) #0.994245) #1.000000;
((EB, NEB) #0.996440,((V, L) #0.240490, (J, (W, PS) #0.859100) #0.949955) #0.955830) #1.000000;

Table B3.17. BPP species estimation (Yang 2015) results for the *Cynocephalus* biparental dataset (101 partitions, with 44,072 sites after removal of sites with missing data, clean data=1).

Model	**Nodes**[a]	**Posterior Probability**	**Posterior Probability**	**Posterior Probability**	**Posterior Probability**
1	00	0.0000	0.0000	0.0000	0.0000
2	10	0.3913	0.2381	0.3365	0.3799
3	11	0.6087	0.7619	0.6635	0.6202
Parameters		Parameters estimated from data			
theta (α, β)		2, 2040	2, 1000	2, 2040	2, 2040
tau (α, β)		2, 771	2, 771	2, 10	2, 1000

[a]The ancestral nodes (left to right) for *Cynocephalus*: **1)** Mindanao to Leyte+Dinagat, **2)** Leyte+Dinagat.

Guide tree with PP for each node (i.e., support for distinct species) in the four analyses shown above (left to right):

(M, (L, D) #0.608665) #1.000000;
(M, (L, D) #0.761880) #1.000000;
(M, (L, D) #0.663515) #1.000000;
(M, (L, D) #0.620150) #1.000000;

APPENDIX C

SUPPLEMENTAL TABLES

Table C3.1

Dermopteran PSGs shared across 4 analyses: 1) Dermopteran branch (7 taxon data set=8,154 genes, *Galeopterus* only), 2) Dermopteran stem branch (8 taxon data set, 4,897 genes, *Galeopterus* + *Cynocephalus*), 3) *Galeopterus* branch (8 taxon data set), 4) *Cynocephalus* branch (8 taxon data set)	
ABCC1	*ASF1A*
ABHD14B	*ASPDH*
ABLIM2	*ATF3*
ABT1	*ATG14*
ACADVL	*ATP10B*
ACKR2	*ATP5D*
ACOT8	*ATP6V1C2*
ADARB1	*AUP1*
ADM	*B4GALT3*
ADORA2B	*BAG1*
AGMAT	*BAIAP2L2*
AGPAT3	*BANF2*
AIFM2	*BARHL1*
AKIP1	*BCL2L12*
ALDH9A1	*BCS1L*
ANAPC5	*BFSP1*
ANKZF1	*BIN2*
APAF1	*BLOC1S4*
APH1A	*BRD8*
APOA1BP	*BTG4*
APOE	*C10orf90*
APOH	*C12orf57*
APOLD1	*C16orf93*
APPL1	*C19orf67*
AQP9	*C1orf116*
ARFGAP1	*C1orf194*
ARHGAP44	*C1orf228*

Table C3.1 Continued.

ARHGEF5	*C1QL4*
C21orf62	*CIITA*
C2orf81	*CLEC4E*
C2orf81	*CLPP*
C6orf222	*CMTM3*
C8orf4	*CMTM5*
CAV3	*COL4A4*
CC2D1A	*COQ10A*
CC2D1B	*COX14*
CCDC108	*CPOX*
CCDC114	*CRACR2B*
CCDC115	*CREB3L4*
CCDC152	*CRTAM*
CCDC155	*CSNK2A2*
CCDC33	*CSRNP1*
CCDC59	*CTC1*
CCL7	*CTNS*
CCR9	*CTRL*
CD300LG	*CX3CR1*
CD40	*CYP17A1*
CD53	*DACT2*
CD5L	*DAD1*
CD72	*DBN1*
CDC7	*DCAF7*
CDH22	*DDX20*
CDHR1	*DDX25*
CDKL1	*DEGS2*
CDPF1	*DENND1C*
CDR2	*DEPTOR*
CDRT4	*DFFB*
CFP	*DGAT1*
CHAF1B	*DMAP1*
CHFR	*DMP1*
CHI3L1	*DNAJC22*
CHRD	*DRAM2*

Table C3.1 Continued.

CHST11	*DUOXA1*
CHST4	*DZANK1*
CIART	*EBPL*
EDN3	*FTSJ2*
EFHD1	*GAN*
EIF4H	*GAREM*
ELANE	*GAS8*
EMC9	*GCHFR*
EMP1	*GCN1L1*
EMP3	*GDAP2*
ENTPD7	*GEMIN6*
EPO	*GHRH*
ERMN	*GIGYF1*
ERP29	*GJA5*
ESAM	*GJB2*
EXOSC3	*GLB1L3*
EXOSC5	*GLE1*
F12	*GLIS1*
F2	*GNG13*
FADD	*GOLGA4*
FAM101B	*GOSR1*
FAM107A	*GPKOW*
FAM159B	*GPR149*
FAM163A	*GPR153*
FAM192A	*GPR50*
FAM196B	*GPR84*
FAM26F	*GPRC5A*
FAM53C	*GPX7*
FANCD2OS	*GRHPR*
FAT4	*GRN*
FBXL15	*HARS*
FBXL16	*HAUS7*
FCGRT	*HDAC7*
FCRLA	*HDDC2*
FEZ2	*HEMGN*

Table C3.1 Continued.

FGF18	*HEMK1*
FGFBP1	*HMCES*
FHAD1	*HMGCR*
FOLR2	*HNRNPA2B1*
FOXI3	*HOXB6*
HOXB9	*LRRC46*
HPCAL4	*MAL*
HSD11B1	*MAP3K4*
HSD17B2	*MARCH3*
HSD17B6	*MARCO*
HTR3A	*MB*
HTRA2	*MBD3L1*
ICAM2	*MC2R*
ICK	*MEGF6*
IFNB1	*MESDC1*
IFT52	*MFAP2*
IL16	*MFSD6*
IL17RB	*MLH3*
IL17RC	*MMD*
IL27	*MMP11*
INHBE	*MMP17*
INIP	*MOCS3*
INPP5J	*MOGS*
INSC	*MPHOSPH6*
INTS5	*MRPL50*
IQCF6	*MRPL51*
IRGC	*MRRF*
ITPKC	*MSI1*
IZUMO1	*MSLN*
JMJD1C	*MSX2*
KIAA0753	*MTRR*
KIAA1614	*MYBPHL*
KLF1	*MYCBPAP*
KLHL38	*MYCT1*
KLHL6	*MYO7A*

Table C3.1 Continued.

LAMTOR2	*NAT10*
LEAP2	*NDEL1*
LGALS12	*NEURL3*
LHX3	*NFE2L2*
LMCD1	*NFIL3*
LMO7	*NKTR*
LRRC14B	*NOSIP*
NOTUM	*POLR1E*
NOXRED1	*PPM1K*
NRSN1	*PROCA1*
NRSN2	*PROP1*
NSMCE4A	*PROSER2*
NT5C	*PRRG3*
NTHL1	*PRSS35*
NUTM1	*PSMB8*
NXPH4	*PSRC1*
OPALIN	*PTGR1*
ORMDL3	*PTRF*
OTOF	*QTRT1*
OVCA2	*RAB20*
P2RX3	*RAB36*
P2RY6	*RABAC1*
PAPPA	*RAD21L1*
PARP14	*RAD51D*
PARP9	*RBFOX2*
PAX2	*RBM22*
PAX8	*RBM34*
PCBP4	*RBM48*
PDCL	*REC8*
PDE1B	*RELB*
PEG3	*RERG*
PGAM5	*RETSAT*
PIF1	*RFWD2*
PIGS	*RFX6*
PIK3CG	*RGAG4*

Table C3.1 Continued.

PIK3IP1	*RHNO1*
PIPOX	*RIBC1*
PKMYT1	*RIBC2*
PKP3	*RNASEH2C*
PLEKHB1	*RNASEL*
PLEKHG6	*RNF114*
PMCH	*RNF168*
PODNL1	*RNMTL1*
POLA2	*RP1L1*
RPAP1	*SOD2*
RPUSD2	*SOHLH1*
RSF1	*SPATA18*
RTN2	*SPATA32*
SAMD10	*SPATS2*
SAMHD1	*SPCS1*
SAP25	*SPPL3*
SATB1	*SPTAN1*
SATB2	*SPTY2D1*
SCAF11	*SPZ1*
SCN3B	*SRD5A2*
SENP2	*SRF*
SENP5	*SRGN*
SERPINA5	*SSH2*
SERPINF1	*SSNA1*
SF3B6	*STAC3*
SFRP1	*STAR*
SHQ1	*STMND1*
SLBP	*STX5*
SLC18A2	*SUOX*
SLC20A1	*SWAP70*
SLC22A18	*SYNC*
SLC24A5	*SYNE4*
SLC25A33	*SYTL3*
SLC25A37	*TAOK3*
SLC2A9	*TATDN3*

Table C3.1 Continued.

SLC35E1	*TTLL1*
SLC35F5	*TWISTNB*
SLC37A3	*TXNIP*
SLC3A2	*UBXN1*
SLC51A	*UBXN11*
SLK	*UNC93B1*
SMARCC1	*USH2A*
SMG7	*USP26*
SMTNL1	*USP43*
SNX20	*VCPKMT*
SNX31	*VGF*
TMEM176A	*VIM*
TMEM177	*VPS11*
TMEM225	*TEX35*
TMEM25	*TGFBR3*
TMEM263	*THEM4*
TMEM74B	*THEMIS2*
TMEM86B	*TIMD4*
TMEM8B	*TIMM21*
TMF1	*TLDC2*
TNFAIP8L2	*TMEM129*
TNFRSF1A	*TMEM139*
TNKS1BP1	*TMEM156*
TNNI1	*TMEM168*
TNNI3	*VSTM2A*
TRH	*VTN*
TRMT2A	*VWA7*
TRMU	*WBP1*
TROAP	*WBP1L*
TRPT1	*WDR13*
TSPAN1	*WDR74*
TSPAN32	*WDR87*
TSPO2	*WDYHV1*
TSR2	*WIBG*
TSSK3	*YAF2*

Table C3.1 Continued.

ZCCHC2
ZCWPW1
ZDHHC23
ZDHHC24
ZHX3
ZNF317
ZNF648
ZNF786
ZYX

Table C3.2 Genes under positive selection contained within enrich sensory modalities for colugos.

Disease Phenotype	#Genes	Colugo PSG Webgestalt Genes	Enrichment Statistics	Additional PS genes not annotated in Webgestalt but associated with similar disease phenotypes
Cardiovascular Diseases	19	*ADM*, APOH*,* ***APOE****,* ***CD40****, CHI3L1*,* ***CX3CR1****,* ***EPO, F2, F12****, GJA5, HMGCR*, ITPKC, MTRR, PAPPA, PSRC1,* ***SERPINF1****, SLC2A9, TNNI1, TNNI3*	C=425;O=19;E=4.61;R=4.12;rawP=2.76e-07;adjP=3.93e-05	
Protein Deficiency	16	*ACADVL,* ***BCS1L****,* ***CAV3****, CFP, CYP17A1,* ***F12, F2****, GAN, IL27,* ***LHX3****, MC2R, PROP1, SPTAN1, SUOX, TNFAIP8L2, UNC93B1*	C=356;O=16;E=3.86;R=4.14;rawP=2.24e-06;adjP=0.0002	
Muscular Atrophy/Muscle weakness	7	*ATP10B,* ***CAV3****, DDX20(GEMIN3), FAM159B, FAM196B, GEMIN6, MB*	C=136;O=7;E=1.48;R=4.74;rawP=0.0007;adjP=0.0086	*EXOSC3, PTRF, STAC3*
Skeletal muscle function/disease	4			*SLC18A2, TNNI1, TNNI3, YAF2*
Hearing Loss, Sensorineural	8	***BCS1L****,* ***F2****, GJB2,* ***LHX3****, MYO7A, OTOF, TRMU, USH2A*	C=140;O=8;E=1.52;R=5.27;rawP=0.0002;adjP=0.0040	*CLPP, COL4A4,* ***FAT4****,* ***LMO7***
Eye Diseases	12	*BFSP1, BLOC1S4,* ***CD40****, CDHR1,* ***CX3CR1****,* ***EPO****, GJB2, MYO7A, PAX2, RP1L1,* ***SERPINF1****, USH2A*	C=368;O=12;E=3.99;R=3.00;rawP=0.0008;adjP=0.0097	*ABCC1, BFSP1, GNG13, INSC, INPP5J, PDCL, VIM*

Table C3.2 Continued.

Retinal Diseases	**10**	*APOE*, ***CDHR1***, ***CTC1***, *CX3CR1*, *EPO*, ***MYO7A***, ***RP1L1***, *SERPINF1*, ***SOD2***, ***USH2A***	**C=247;O=10;E=2.68;R=3.73;rawP=0.0004;adjP=0.0058**	***ADM**, **HPCAL4**, **HPCAL5***
Macular Degeneration	8	***APOE***, *CDHR1*, *CFP*, ***CX3CR1***, *RP1L1*, ***SERPINF1***, *SOD2*, *VTN*	C=112;O=8;E=1.22;R=6.58;rawP=3.22e-05;adjP=0.0011	
Deaf-Blind Disorders	4	*CDHR1*, *MYO7A*, *OTOF*, *USH2A*	C=23;O=4;E=0.25;R=16.03;rawP=0.0001;adjP=0.0025	
Joint/digital/skeletal deformities	8			*ABT1*, *CHRD*, *DMP1*, ***FAT4***, ***LMO7***, *MARCO*, *PSMB8*

*genes that have altered expression in the retina of aging and diabetic persons boldface indicates genes with pleiotropic effects.

Table C3.3. Positively selected genes on the primate ancestral branch.

PSGs on the Primate Ancestral branch, 7 taxon dataset, 194 genes	
A4GALT	*CDHR2*
ACSBG1	*CEBPZ*
ACTRT3	*CFAP43*
ADAL	*CHPF2*
ADCY9	*CHST3*
ADNP2	*CNPY4*
AHCTF1	*COL9A3*
AKAP8L	*CORO2A*
ALG12	*CPD*
ALKBH2	*CRISPLD2*
ALOXE3	*CSN2*
ANO9	*CSRP2BP*
AP5Z1	*CTSB*
APOA2	*CYP2W1*
ARFRP1	*DAGLB*
ARHGEF16	*DBH*
ATP10B	*DCAKD*
ATP7B	*DCLRE1B*
ATXN10	*DEFB129*
AUNIP	*DHCR24*
B4GALT6	*DIABLO*
BPIFB1	*DLL1*
C2orf71	*DNM1*
C2orf81	*DNMT1*
C5	*DOC2A*
C6orf58	*DPAGT1*
C8orf37	*DYRK2*
C9orf9	*E2F6*
CA14	*EEF2*
CAMK2A	*EID2B*
CCDC186	*EIF2AK2*
CCDC73	*ELAVL3*
CCDC77	*ELMO3*
CD19	*ENOX1*
CD38	*EPHB6*
CD86	*ERMN*

Table C3.3 Continued.

EVPL	*MFGE8*
F2	*MORN1*
FAM26D	*MRPL46*
FETUB	*MRPL55*
FGB	*MS4A7*
FMO1	*NCS1*
GALNT16	*NDUFV3*
GCN1L1	*NEFM*
GDF2	*NME8*
GLYAT	*NOL11*
GLYATL3	*NOSTRIN*
GPRC5D	*NPHP1*
GPS1	*NPHS1*
GRAMD1B	*NXNL1*
HELZ	*ODF2L*
HFM1	*OR4C5*
HIVEP2	*OXNAD1*
HJURP	*P4HTM*
HSH2D	*PCP2*
HSPG2	*PCTP*
IFI35	*PIP5K1C*
ITGAL	*PITRM1*
KCNV2	*PKN3*
KIAA1161	*POLR3A*
KIAA1551	*PRR14L*
KLF6	*PVRL2*
KRT36	*QTRTD1*
KRT40	*RANBP3L*
LALBA	*RAP1GAP2*
LAMC3	*RASSF7*
LIPF	*RBM15*
LONP1	*RGS3*
LPXN	*RNASEL*
LRCH3	*S100A9*
LRFN3	*SACS*
MAEA	*SETDB1*
MAP1A	*SGMS2*

Table C3.3 Continued.

MCM4	*WDR87*
MDM4	*WFS1*
SHQ1	*YIPF2*
SLC16A10	*ZC3H12A*
SLC3A2	*SHBG*
SLC41A1	*SHE*
SLC52A3	*ZCCHC6*
SLC6A3	*ZDHHC7*
SLC6A6	*ZNF395*
SLC7A2	*ZNF541*
SNX20	*ZWILCH*
SPP1	
SPR	
SPTBN2	
SSC5D	
ST6GAL2	
SUV39H1	
TDRD6	
THSD7A	
TICRR	
TIMD4	
TK2	
TLR8	
TMCC2	
TMEM139	
TMEM17	
TMPRSS2	
TNC	
TOLLIP	
TOM1L2	
TOMM40L	
TOP3A	
UQCRC1	
USP25	
UVSSA	
VTN	
WBP1	

Table C3.4. Genes under positive selection and their known functions placed in generalized functional categories for ancestral Primates.

Primate Ancestor PSG Webgestalt results	
KEGG pathway	Metabolic pathways
C=1130; O=19; E=5.00; R=3.80; rawP=7.81e-07; adjP=2.81e-05	
A4GALT	alpha 1,4-galactosyltransferase
ALG12	asparagine-linked glycosylation 12, alpha-1,6-mannosyltransferase homolog (S. cerevisiae)
B4GALT6	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 6
CD38	CD38 molecule
CHPF2	chondroitin polymerizing factor 2
DBH	dopamine beta-hydroxylase (dopamine beta-monooxygenase)
DHCR24	24-dehydrocholesterol reductase
DNMT1	DNA (cytosine-5-)-methyltransferase 1
DPAGT1	dolichyl-phosphate (UDP-N-acetylglucosamine) N-acetylglucosaminephosphotransferase 1 (GlcNAc-1-P transferase)
LALBA	lactalbumin, alpha-
LIPF	lipase, gastric
NDUFV3	NADH dehydrogenase (ubiquinone) flavoprotein 3, 10kDa
PIP5K1C	phosphatidylinositol-4-phosphate 5-kinase, type I, gamma
POLR3A	polymerase (RNA) III (DNA directed) polypeptide A, 155kDa
SGMS2	sphingomyelin synthase 2
SPR	sepiapterin reductase (7,8-dihydrobiopterin:NADP+ oxidoreductase)
ST6GAL2	ST6 beta-galactosamide alpha-2,6-sialyltranferase 2
TK2	thymidine kinase 2, mitochondrial
UQCRC1	ubiquinol-cytochrome c reductase core protein I

Table C3.4 Continued.

KEGG pathway	**ECM-receptor interaction**
C=85; O=5; E=0.38; R=13.28; rawP=3.98e-05; adjP=0.0007	
HSPG2	heparan sulfate proteoglycan 2
LAMC3	laminin, gamma 3
SPP1	secreted phosphoprotein 1
TNC	tenascin C
VTN	vitronectin
disease	**Brain Diseases**
C=411; O=12; E=1.82; R=6.59; rawP=3.58e-07; adjP=0.0001	
ACSBG1	acyl-CoA synthetase bubblegum family member 1
ATP7B	ATPase, Cu++ transporting, beta polypeptide
ATXN10	ataxin 10
DBH	dopamine beta-hydroxylase (dopamine beta-monooxygenase)
DHCR24	24-dehydrocholesterol reductase
F2	coagulation factor II (thrombin)
FGB	fibrinogen beta chain
SACS	spastic ataxia of Charlevoix-Saguenay (sacsin)
SLC41A1	solute carrier family 41, member 1
SLC6A3	solute carrier family 6 (neurotransmitter transporter, dopamine), member 3
SPR	sepiapterin reductase (7,8-dihydrobiopterin:NADP+ oxidoreductase)
SPTBN2	spectrin, beta, non-erythrocytic 2

Table C3.4 Continued.

disease	**Neurodegenerative Diseases**
C=404; O=11; E=1.79; R=6.15; rawP=2.18e-06; adjP=0.0004	
ATP7B	ATPase, Cu++ transporting, beta polypeptide
ATXN10	ataxin 10
CAMK2A	calcium/calmodulin-dependent protein kinase II alpha
DHCR24	24-dehydrocholesterol reductase
NEFM	neurofilament, medium polypeptide
SACS	spastic ataxia of Charlevoix-Saguenay (sacsin)
SLC41A1	solute carrier family 41, member 1
SLC6A3	solute carrier family 6 (neurotransmitter transporter, dopamine), member 3
SPTBN2	spectrin, beta, non-erythrocytic 2
UVSSA	UV-stimulated scaffold protein A
WFS1	Wolfram syndrome 1 (wolframin)
disease	**Schizophrenia**
C=360; O=8; E=1.59; R=5.02; rawP=0.0002; adjP=0.0088	
CAMK2A	calcium/calmodulin-dependent protein kinase II alpha
DBH	dopamine beta-hydroxylase (dopamine beta-monooxygenase)
DOC2A	double C2-like domains, alpha
ERMN	ermin, ERM-like protein
MAP1A	microtubule-associated protein 1A
NCS1	neuronal calcium sensor 1
NEFM	neurofilament, medium polypeptide
SLC6A3	solute carrier family 6 (neurotransmitter transporter, dopamine), member 3

Table C3.4 Continued.

disease	**Bipolar Disorder**
C=344; O=8; E=1.52; R=5.25; rawP=0.0002; adjP=0.0088	
ADCY9	adenylate cyclase 9
CAMK2A	calcium/calmodulin-dependent protein kinase II alpha
DBH	dopamine beta-hydroxylase (dopamine beta-monooxygenase)
GDF2	growth differentiation factor 2
NCS1	neuronal calcium sensor 1
SLC6A3	solute carrier family 6 (neurotransmitter transporter, dopamine), member 3
WFS1	Wolfram syndrome 1 (wolframin)
disease	**Cystinuria**
C=35; O=4; E=0.16; R=25.81; rawP=1.75e-05; adjP=0.0014	
SLC16A10	solute carrier family 16, member 10 (aromatic amino acid transporter)
SLC3A2	solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2
SLC6A6	solute carrier family 6 (neurotransmitter transporter, taurine), member 6
SLC7A2	solute carrier family 7 (cationic amino acid transporter, y+ system), member 2
disease	**Musculoskeletal Diseases**
C=462; O=10; E=2.05; R=4.89; rawP=4.52e-05;adjP=0.0030	
C5	complement component 5
COL9A3	collagen, type IX, alpha 3
CRISPLD2	cysteine-rich secretory protein LCCL domain containing 2
NME8	*NME/NM23 family member 8*
S100A9	S100 calcium binding protein A9

Table C3.4 Continued.

SACS	spastic ataxia of Charlevoix-Saguenay (sacsin)
SPP1	secreted phosphoprotein 1
TK2	thymidine kinase 2, mitochondrial
TNC	tenascin C
UVSSA	UV-stimulated scaffold protein A

*Function/Disease Association (from GeneCards Database (http://www.genecards.org/)

Table C3.5. Colugo sample information for *Cynocephalus volans* (CVO) and *Galeopterus variegatus* (GVA).

Lab ID	Museum	Museum ID	Sex	Collection Date	General Location	Tissue Code	Coordinates	Location
CVO 02	FMNH		M		PHI-LEY	MUS		
CVO 03	AMNH	24958	M(JUV)	1905	PHI-SAM	SKN, MUS	11°55'56.85"N, 125° 2'19.74"E	Samar Island, Philippines
CVO 06	AMNH	24981	M(JUV)	1905	PHI-SAM	MUS	11°54'37.43"N, 125° 2'48.96"E	Samar Island, Philippines
CVO 07	AMNH	85042	Unk	Unk	PHI	SKN-PAT	9.937748°, 124.192405°	Philippines (infered by phylogeny, Pulau Bohol)
CVO 08	AMNH	203257	F	1962	PHI-MIN	SKN-PAT	6°20'18.58"N, 124°58'21.76"E	Tupi, Mindanao
CVO 09	AMNH	16219	F	Unk	PHI-BOH	SKN-PAT	9°47'47.78"N, 124°14'31.78"E	Bohol, Philippines
CVO 10	AMNH	203258	M	1962	PHI-MIN	SKN-PAT	6°20'18.58"N, 124°58'21.76"E	Tupi, Mindanao
CVO 13	AMNH	187860	M	1961	PHI-LEY	MUS, SKN-FOT	11° 0'45.29"N, 124°48'51.83"E	Mount Lobi, Burauen, Leyte Island
CVO 15	AMNH	187861	Unk	1961	PHI-LEY	HAR, SKN	11° 2'38.91"N, 124°48'54.86"E	Mount Lobi, Burauen, Leyte Island
CVO 16	AMNH	207472	F	1960	PHI-MIN	BON-TUR	8° 7'48.82"N, 125° 7'42.49"E	Malayablay, Bukidon, Mindanao Island
CVO 17	NMNH	219289	Unk	23 Jul 1918	PHI-MIN	CRS-SKL	8°20'13.98"N, 123°42'26.42"E	Misamis Occidental, Mindanao, Philippines (Province Northern Mindanao)
CVO 18	NMNH	144659	M	3 Feb 1906	PHI-BAS	CRS-SKL	6°42'10.51"N, 121°58'8.47"E	Isabela City, Basilan Island, Philippines
CVO 19	NMNH	578084	M	10 Apr 1987	PHI-LEY	CRS-SKL, BON	11°33'46.28"N, 124°23'47.50"E	Naval, Leyte Island, Philippines
CVO 21	NMNH	144663	M	17 May 1906	PHI-BAS	MUS	6°42'6.24"N, 121°58'20.73"E	Basilan, Island, Philippines, Isabella City
CVO 22	NMNH	113493	M	-- Sep 1901	PHI-MIN	SKN	8°13'40.88"N, 124°14'42.87"E	Iligan City, Mindanao, Philippines

Table C3.5 Continued.

CVO 24	**NMNH**	**462160**	**F**	**8 May 1975**	**PHI-DIN**	**BON-TUR**	**10°21'12.18"N, 125°36'59.80"E**	**Loreto, Dinagat Island**
GVA 01	CMNH	87909	M		THA		unknown	Surat Thani, Thailand
GVA 03	RMBR	ZRC.4.8112	M		IDN-PEN		unknown	MacRitchie, Singapore, Peninsular Malaysia
GVA 04	KMNH		M		IDN-JAV			Sumur, W. Java
GVA 07	FMNH	171074: Locus 010187			LAO		unknown	Laos
GVA 08	NMNH	154600	F	25 May 1909	IDN-JAV	CRS-NAS	6°45'S, 106°41'E	Mount Salak, W. Java
GVA 09	NMNH	155363	Unk	6 Apr 1909	IDN-JAV	CAR	ca 8 deg S, ca 113 deg E	Jawa Timur Province, E. Java
GVA 10	NMNH	307553	Unk	28 Sep 1957	MYS-PEN	CAR	4.52 deg N, 101.38 deg E	
GVA 11	NMNH	311297	M	17 Jul 1958	MYS	CRS-SKL	6°19'48"N, 99°43'43"E	Kedah, Langkawi Island, Malaysia
GVA 12	NMNH	197203	F	2 Jul 1913	IDN-BOR	CAR	1.23 deg N, 118.73 deg E	Labuan Klambu, Borneo, Indonesia
GVA 13	NMNH	317119	Unk	23 Sep 1960	MYS-BOR	CRS-NAS	5°57'8"N, 116°39'52"E	Ranau, Sabah, Borneo, Malaysia
GVA 14	NMNH	356666	F	8 Feb 1963	THA	CRS-NAS	9.6167° N, 98.55° E	Tambon Muang Klang, Ranong, Amphoe Kapoe, Muang Klang, Thailand
GVA 16	NMNH	104600	M	7 Jul 1900	IDN-NAT	CRS-SKL	2°31'13"N, 109°2'51"E	Sirhassen, Natuna Island
GVA 17	NMNH	115605	F	20 Aug 1902	IDN-SUM	CRS-SKL	1°1'31"N, 104°27'44"E	Pulau Bintan, Rhio Archipelago, Sumatra
GVA 18	NMNH	121749	M	12 Feb 1903	IDN-SUM	CRS-SKL	0°28'35.79"S, 98°23'32.39"E	Pulau Tanahbala, Batu Island, Utara Province, Sumatra
GVA 19	NMNH	143327	F	12 Mar 1906	IDN-SUM	CRS-SKL	1°52'32"N, 101°34'48"E	Pulau Rupat, Sumatra

Table C3.5 Continued.

GVA 21	**AMNH**	**106627**	**F**	**1936**	**IDN-SUM**	**CRS-SKL**	**0°44'23.52"S, 100°48'0.45"E**	**West Sumatra**
GVA 22	AMNH	102703	F(YA)	1934	IDN-SUM	CRS-SKL	2°59'27.77"S, 104°45'24.39"E	Macarah Doewa, Palembang, Sumatra
GVA 23	AMNH	106629	M	1936	IDN-SUM	CRS-SKL	3°34'57.47"S, 102°21'0.06"E	Bengkulu, Sumatra
GVA 26	AMNH	103735	F	1935	MYS-BOR	MUS-FOT	??? 5°47'2.76"N, 117°15'26.89"E	Badang, NE:Borneo
GVA 27	AMNH	107136	F(SA)	1937	BOR	MUS-FOT	0° 0'40.89"N, 109°16'26.80"E	Perboewa (Perbuah)???, Landak, NW: Borneo
GVA 28	AMNH	102704	M(JUV)	1934	IDN-SUM	MUS-FOT	2°59'27.99"S, 104°45'24.24"E	Palembang, Sumatra
GVA 32	AMNH	37202	F(Inf)	Unk	THA	SKN	7.172756°, 99.906227°	Thailand
GVA 35	AMNH	103734	M	1935	MYS-BOR	SKN-PAT	??? 5°47'2.76"N, 117°15'26.89"E	Badang, NE:Borneo
GVA 36	AMNH	54964	M(JUV)	1928	MYN	SKN-PAT	10°16'19.45"N, 98°35'52.55"E	Maliwan, Myanmar (Also known as Burma)
GVA 38	AMNH	101500	M	1929	IDN-JAV	SKN-FOT	6°43'0.00"S, 108°34'0.00"E	Cirebon, Java
GVA 39	AMNH	32644	M(JUV)	1910	MYS-BOR	SKN-FOT, MUS-FOT	1°31'37.79"N, 110°20'40.09"E	Kuching,Borneo (Slightly South), Malaysia-Sarawak
GVA 40	AMNH	101501	F	Unk	IDN-JAV	CAR	6°43'0.00"S, 108°34'0.00"E	Cirebon, Java
GVA 45	NMNH	300827	M	24 Jul 1953	MYS-BOR	CRS-SKL	5°56'13.38"N, 116°40'10.15"E	Ranau, Sabah, Malaysia
GVA 46	NMNH	176431	M	9 Feb 1913	IDN-BOR	CRS-SKL	1.476421°, 118.160730°	Gunung (Mountain) Talisaian, East Kalimantan, E Borneo
GVA 49	NMNH	151887	M	31 Dec 1907	IDN-BOR	CRS	3°30'35.26"S, 116° 9'19.46"E	Pulau Sebuko, Borneo

Table C3.5 Continued.

GVA 51	**NMNH**	**253411**	**F**	**9 Jul 1928**	**THA**	**MAN**	**8°26'2.43"N, 99°46'51.00"E**	**Ban Kiriwong Distribution Company, Thailand**
GVA 53	NMNH	115606	F	20 Aug 1902	IDN-SUM	CAR	1° 8'12.13"N, 104°25'32.71"E	Bintang Island
GVA 55	NMNH	112427	M	5 Jun 1901	MYS	CRS-NAS	2°26'38.95"N, 104°31'29.09"E	Johor, Aur Island (Pulau)
GVA 56	NMNH	121854	M	4 Mar 1903	IDN-SUM	BON-TUR	0° 7'45.41"N, 98°40'55.43"E	Pini Island (Pulau)
GVA 57	NMNH	123035	M	14 Jun 1903	IDN-SUM	BON-TUR, CRS-NAS	0°18'16.91"S, 103°35'19.03"E	Pulau Bakong, Indonesia
GVA 58	NMNH	125124	F	1 Sep 1904	IDN-BOR	CRS-NAS	1°36'15.65"S, 108°52'44.09"E	Pulau Karimata
GVA 61	NMNH	153864	M	29 Aug 1908	IDN-BOR	BON-TUR	2°32'41.37"S, 110°12'17.35"E	Parui, Sungai Kendawangan, Borneo
GVA 62	CVHM	T114		Unk	NAM	SKN-PAT	unknown	Vietnam
GVA 63	CVHM	T203		Unk	NAM	SKN-PAT	unknown	Vietnam
GVA 64	RMBR	ZRC.4.875	F	11550	IDN-NAT	CRS-SKL	2°30'1.20"N, 109° 5'16.05"E	Pulau Serasan, Natuna Isls.
GVA 69	RMBR	ZRC.4.881	M	**-?Sep-1925	IDN-NAT	CRS-NAS	3° 8'28.27"N, 106° 5'14.68"E	Pulau Siantan, Anambas Isls.
GVA 71	RMBR	ZRC.4.870	F	**-Jun-1894	IDN-NAT	CRS-NAS	3°56'1.16"N, 108°12'34.91"E	Pulau Natuna-Besar, Bunguran Barat
GVA 76	RMBR	ZRC.4.882	F	9392	IDN-NAT	CRS-NAS	3° 8'28.27"N, 106° 5'14.68"E	Pulau Siantan, Anambas Isls.
GVA 99	RMBR	ZRC.4.946	M	**-***-1935	IDN-SUM	CRS-MAN	3°25'55.89"N, 98°44'10.30"E	Deli Sempang Regency, NE Sumatra
GVA 115	RMBR	ZRC.4.986	M(JUV)	11541	IDN-NAT	CRS-NAS, BON-TUR	2°54'54.29"N, 108°50'37.74"E	Pulau Subi-Besar
GVA 125	NMNH	197202	F	2 Jul 1913	IDN-BOR	CRS-SKL	1°15'31.59"N, 118°41'52.08"E	Labuan Klambu
GVA 128	NMNH	123086	F	2 Aug 1903	IDN-SUM		0°19'33.30"S, 104°28'0.76"E	Pulau Penuba, Indonesia

Table C3.5 Continued.

GVA 129	**NMNH**	**267397**	**Unk**	**-- --- ----**	**IDN-SUM**		**2°57'29.71"N, 99° 3'52.81"E**	**Pematangsiantan, Indonesia => Siantan, Indon.**
GVA 133	NMNH	121748	M	6 Feb 1903	IDN-SUM		0°28'6.55"S, 98°23'32.79"E	Batu Islands, Pulau Tanahbala
GVA 134	NMNH	121853	M	2 Mar 1903	IDN-SUM		0° 6'59.26"N, 98°40'55.36"E	Batu Islands, Pulau Pini
GVA 138	NMNH	104447	M	15 Dec 1899	THA		6°32'37.07"N, 99°18'27.24"E	District Satun, Thailand, Ko Adang
GVA 139	NMNH	083276	M	1 Mar 1896	THA		7°14'40.72"N, 99°23'16.65"E	Ko Ta Li Bong
GVA 140	NMNH	084421	F	4 Mar 1897	THA		7° 8'1.52"N, 99°42'1.55"E	Ban Lao Trong, Trang, Thailand
AF460846	Genbank						unknown	W. Java, Inferred
JN800721	Genbank						09°53'51.6372"N, 098°43'16.6188"E	"Malayan", Inferred to be Thailand
AJ428849	Genbank						01°12'49.2300"N, 110°25'46.1028"E	Inferred to be NW: Borneo (Possibly Sarawak)

Table C3.5 Continued.

Legend						
NMNH = National Museum of Natural History (Smithsonian)					IDN = Indonesia	MUS = Muscle
AMNH = American Museum of Natural History					THA = Thailand	CRS = Crusties
RMBR = Raffles Museum of Biodiversity Research					MYS = Malaysia	SKN = Skin
FMNH = Field Museum of Natural History					SUM = Sumatra	BON = Bone
CMNH=Museum of Texas Tech University					BOR = Borneo	NAS = Nasal
KMNH=Kitakyushu Museum of Natural History					JAV = Java	SKL = Skull
					NAT = Natuna	MAN = Mandible
M = Male (Adult)					MYN = Myanmar	TUR = Turbinate
F = Female (Adult)					PHI = Philippines	PAT = Patagium
YA = Young Adult					MIN = Mindanao	FOT = Foot
SA = Sub-Adult					LEY = Leyte	KER = Keratin
Unk = Unkown					BOH = Bohol	
JUV = Juvenille					SAM = Samar	
IMM = Immature					BAS = Basilan	
					DIN = Dinagat	
					N/A = Not Available	

Table C3.6. Primers used to amplify probing sequences used in capture hybridization reactions.

#ID	Oligo_name	GVA_04 Contig Name	Forward_Primer	Reverse_Primer
1	ASIP_01_F	NW_007732451.1	TGTGAGTGAATGAGGCAGGA	GGACACTTTATTTCCCCAGGA
2	ASIP_02_F	NW_007732451.1	ACACCAAGCCCATTTTCAGA	GCAAAAGAATGAGTGATGGTGA
3	ASIP_03_F	NW_007732451.1	AAAGCTGTGGAGCTGAGTGG	GGGATGGCGAGAGCTACTTA
4	HMGA2_02_F	NW_007726460.1	TGCATCTCCAAAAGGAAGTG	CATGAACTTTTTATTCTAGGCATGG
5	HMGA2_03_F	NW_007726460.1	TTACTCCAGGGGAGCCTTTT	CATTTCACCTGAATTTTATCACCA
6	HMGA2_04_F	NW_007726460.1	CCCTTGAACCTTAGAGAAACCA	GACCCAAGAGTAGTTCAATCTGC
7	HMGA2_05_F	NW_007726460.1	AACTGATGACTGGCATGCTG	AGTGGAAAGACCATGGCAAC
8	IGF1_01_F	NW_007729881.1	TTTTAATGTCTGCTAACCCTGTCA	AACATCTGCACCTGCGAAA
9	IGF1_02_F	NW_007729881.1	CATCTTTCATGTACATAGTCGATGTTT	CCTTTGAGCGAAGTTCACCT
10	IGF1_03_F	NW_007729881.1	TGTTCATAGCATCCTCCTAAGTCA	TTTTCCACCATCTTCCCATT
11	IGF1_04_F	NW_007729881.1	GGCTTGTCCATCTACATATGTCCT	TGGTAGGGACAGACAAACAAGA
12	LCORL_02_F	NW_007726234.1	CATCACTGAGACAAAAACAAGC	AATTTTTGATAGTCATTGCGTGA
13	LCORL_03_F	NW_007726234.1	CCTAAAAGACTTCAAGAATAAGTGAAA	TTTTCTTGGGCATCATTCTATTT
14	LCORL_04_F	NW_007726234.1	GAAAATATCTGACTGTACCTAATTCCA	TGATTCTTCACAGTCAACACCA
15	LCORL_06_F	NW_007726234.1	GAAATTACATATTTCCCTTATAATTGC	CAAATACATTGGTTATAGAGTCAC
16	LCORL_07_F	NW_007726234.1	CGAAAACATGCTATTTGCTGA	CCTGCCAGGATTATTTAACTTTT
17	OPN1MLW_01_F	NW_007727788.1	CTGCAGGAGGCTCCAGTTC	CTGGGTGTTTGCCTCCTTT
18	OPN1MLW_02_F	NW_007727788.1	GCCAGCCCCTCTCTCTACTC	GCACTGTGTGGCCACTATCT
19	OPN1MLW_04_F	NW_007727788.1	CCTACCCAAGTCAGTGAGTGC	AGGAGGGTGGAAAGTTGGTC
20	OPN1MLW_05_F	NW_007727788.1	TTACCCATGTGAAGCTGCTG	CAACGTGCCAGCTAACGAT
21	OPN1MLW_06_F	NW_007727788.1	CCCATGTGCAGGGAATATCT	GATTTGTGAGCCTGCTCCTC
22	OPN1SW_01_F	NW_007729661.1	CAGCCTGTCACCAGACCTGT	TGAATATGGGTCAGCCCACT
23	OPN1SW_03_F	NW_007729661.1	CTCCACAATCGCCTCTCTTC	CCTTTGAGCGCTACCTTGTC
24	OPN1SW_04_F	NW_007729661.1	TCTCCAAGTGCTGTGAGTGG	AGGAAGTGGGGGTAGAAGGA

Table C3.6 Continued.

25	ZFAT_02_F	NW_007728386.1	CTCTTCTACCAAGCAGGCACT	AACAGTGATAATGCTTTAAAATAG
26	ZFAT_03_F	NW_007728386.1	CTCCCAGAACATACACTCACCA	CATTAGGAGCGCGACAAACT
27	ZFAT_04_F	NW_007728386.1	AAATACGGACTTAAAAATGAAAGATT	TACTTTGCATGGGTGTCCTG
28	ZFAT_05.1_F	NW_007728386.1	CCAAGTGCTAACTGCAACACA	GCCCAGGAGACTTTGGTCTT
29	ZFAT_05.2_F	NW_007728386.1	ATCCTCAGGCTGGGCATTC	TGAATGCCTCCATTTATCCTG
30	ZFAT_07_F	NW_007728386.1	AGGCTCTGGGGGAGAACTTA	GAGCCCAGATTTAAGCGAGA
31	ZFAT_08_F	NW_007728386.1	TCTTAGCTGGCACACTTCTCAG	ACATCCTGGTTTAACTTTTTAGGC
32	ZFAT_09_F	NW_007743377.1	GGCAGAATCCACTTCTTTTCA	CAGCTTCAGTTGGACTTGGA
33	ZFAT_10_F	NW_007731851.1	TCTCCAGCTCGTACAACAGC	CTCCTTAGTGAGGTGCTCCAG
34	ZFAT_11_F	NW_007731851.1	AGGAAGACAGTTTTTCAAAGGAA	CTTGTGTAGGCAGGTGCTTG
35	ZFAT_12_F	NW_007731851.1	AGAAGGCAGGTAGGAGCACA	AAATCTCCGCCCTGAAAAAT
36	ZFAT_13_F	NW_007731851.1	TTAAAGCTCCACCTGTGCTG	TGGAAGGGGCAGATTAGAGA
37	ZFAT_14_F	NW_007731851.1	ATCACCTCAGTGTCCCGTGT	AGTGATCACCGAGAGCCTGT
38	ZFAT_15_F	NW_007731851.1	GGGAAGCATAAGCAAAAGCA	GTGACCCTATGATGGGGATG
39	MC1R_Gene01_F	NW_007726285.1	GTAAGCTACCCCCTCCTGCT	AGCAGAGAAGCACCTCCTTG
40	MC1R_Up02_F	NW_007726285.1	CATTAGTGGCTGTCGGGTCT	CCATTTCTTGGGTGGACTTG
41	X_Contig69748_F	NW_007728932.1	GACAGCATGGTGAAAGCTGA	CACAGACACCTTCTGCCTGA
42	X_Contig94865_F	NW_007731575.1	TGGCAACTGCTTCTTGTGTC	TACAGCACAGGCAGAATGGA
43	X_Contig107949_F	NW_007732021.1	TCATCCCGCTTCATTAGTCA	GAAAAAGCATGGGGAAAACA
44	X_Contig14823_F	NW_007729443.1	AGCAGACTGTCCCAGGAATG	CAGTCTGTGGTGAGCTCGAA
45	X_Contig45101_F	NW_007726692.1	GGACCTACAGAGCCTAGGGAAT	TTTCAAAACAGTAAAGCATTGTGA
46	X_Contig20894_F	NW_007726883.1	TGGCTGTCCTGGTAACACAA	CGGAACATTCAAAGTGCTGA
47	X_Contig18190_F	NW_007731448.1	CCAGTGACAGCCCAAACC	TGGTGATTGTGTCCCTGAGA
48	X_Contig133814_F	NW_007742817.1	CGTGTGAGAAAACCAGTGGA	TTCCTCAAACCAACACTGGA
49	X_Contig49331_F	NW_007728345.1	CAATCTGTGCGGTTTTCTCA	TGCAGGGTGTGCACATTTAT

Table C.6 Continued.

50	X_Contig37025_F	NW_007728106.1	CCCTTCAGGTCCCAAGGTAT	GCCTCTTCTGCACTCTTTGG
51	X_Contig20163_F	NW_007726799.1	AGGTCCCTGTGGTGTTTCTG	TCCTGGAAGAAAGGTTCCAA
52	X_Contig62835_F	NW_007737834.1	TGGAGAAAAGTGGGAAATGG	CAATGAGAATTGGGGGAAGA
53	X_Contig10345_F	NW_007727512.1	TTGCAGTTTCATTCATTGCAG	CCACTCTGGCCTCCCTCTAT
54	X_Contig5456_F	NW_007726603.1	ACTGATCACCTGGCTGCTTT	CCTGACCGTCACCTATAGCC
55	X_Contig13029_F	NW_007726216.1	TGGCACAGACAGGTAAATGC	GGGGTATGGTGTTTCACAGG
56	X_Contig26106_F	NW_007738000.1	GGGGACAGGAGAGAGAGCTT	TCCTACGTTACCCGATTCCA
57	X_Contig31019_F	NW_007730962.1	CACAGTGTCCGCTAGTGCTC	TTCTTCCGAGGAATCACTGC
58	X_Contig49517_F	NW_007726878.1	TTTCTCCAGGTAAATGGTGATCT	GTGTTTTGTTCCAAGCCACA
59	X_Contig68509_F	NW_007728204.1	CCTTTCAAATGGCTCTCAGC	TCAGCCAGAGATCATCAGCA
60	X_Contig1374_F	NW_007726992.1	CCTCCATCTGTATCCCCAGA	AAGGCCCAACATGACTTCAG
61	X_Contig152258_F	NW_007734669.1	GAGGCAGAGGAGGTCAAGC	GCACAGCACACCCTTATGTG
62	X_Contig20952_F	NW_007726962.1	TCTGAAAAGGAGGGTGATGC	CCATTCAGCTTGTTGGGAGT
63	X_Contig27453_F	NW_007727052.1	CCTCTTGGAGAAATGGCAAG	TGGCAAAGAGAGATGGAAGG
64	X_Contig110324_F	NW_007731456.1	AATGCAAATGACAACAAAGAGC	AAAAGATCACCTCCCCTCCA
65	X_Contig254220_F	NW_007733235.1	ATGGAGACTTCCATGGCTTG	ACAGGATGTCCAGTCCTCCA
66	X_Contig17617_F	NW_007727244.1	GTGATGAGTCCCGACATTCC	TTGCTCTGTGGGTAGACACG
67	X_Contig35386_F	NW_007728720.1	CAACTGGCATGTTCCACAGT	TCTGGACACGTCAGATTTGC
68	X_Contig25324_F	NW_007734669.1	AGTGTCTGCCCAGTGAAACC	CTGCCAAATTCCCTGAAGAG
69	X_Contig46525_F	NW_007727198.1	CGGGAGAAGAAACAGAATCG	CAAGACCTTGGCCTGGAGTA
70	X_Contig3063_F	NW_007732874.1	CAAGAAAGTCGGCTGGAGTC	GCCACTGCCATTCTTAGGAG
71	X_Contig39358_F	NW_007726833.1	GGCAGCAAGGGCTGTAATAG	CCTGGCTTCCCCTTTCTTTA
72	X_Contig271_F	NW_007727644.1	GGGGCAAGTGCCATATATTC	AGCAGGTGTGATCCAGAAGAA
73	X_Contig17300_F	NW_007728011.1	TGGCAGCAGCAAGTTCATAG	AGGACAACCCCTTTTCCTTG
74	X_Contig19627_F	NW_007728768.1	AACTTCGACCAACTTACCCATT	TGGGTGTAGGGGAAGAGAAA

Table C.6 Continued.

75	X_Contig8739_F	NW_007726121.1	GGAATGACTCTGGGCTGCTA	GGCTCCAGAAACTGCAGAAC
76	X_Contig2559_F	NW_007728914.1	TCTGAGTGATTTAGGCCTGCT	GACCAGGCAAGTCCAAACAT
77	X_Contig55418_F	NW_007730191.1	GCAGTGGGATGTGACAACTG	TCCAGCAGTGGAAGAACCTT
78	X_Contig8081_F	NW_007730557.1	TCTCACACTGCAGCCTTCAC	ACGGTGTGATGGAAAAGGAA
79	X_Contig65360_F	NW_007735886.1	TGACCCAACTGGGAATAACC	CTGGTCCTCTTGGGCTGAAT
80	X_Contig25393_F	NW_007727647.1	CCCAGGTTGGACGTAATGAT	GGTCATTTTCCTCTGGGCTA
81	X_Contig22852_F	NW_007726384.1	AAGTTTGCAAGTCAGCCTCAA	TATGGTCCATGGCTCCTCAT
82	X_Contig40258_F	NW_007727349.1	CTCCAGCTTGTTGGTGATGA	ATGAGCGAGCAGAAGAGAGC
83	X_Contig65408_F	NW_007726476.1	AGACCCGAAAGCAGCCTTTA	TTGTATGGCACAGTGGAGGA
84	X_Contig14915_F	NW_007755616.1	GATGGAAGTGCTGGAGGATG	TGACCTGCTCTTGCTCCTTT
85	X_Contig268296_F	NW_007732238.1	TCTGGAGGTTCTTTGCTGCT	GGATGAAAATGCAAGTTGGAC
86	X_Contig21647_F	NW_007730013.1	AACGCATGGTAGAAGCATCA	TTAGCGAGGGAATGCTTGTC
87	X_Contig5180_F	NW_007726558.1	AGATGGCACAGCTGGGATAC	TCACTTGGAGCTGCTTGGTA
88	X_Contig15338_F	NW_007731060.1	TGCTGTCTTTGCTCAGTGCT	TACAGGCCAACCAAGCTTCT
89	X_Contig17341_F	NW_007730327.1	CCCATGATTTCTGCCCTCTA	TTGTTTTAGGGCCACATGGT
90	X_Contig12545_F	NW_007731408.1	CAGGTGAGGTTGGTTTCAGG	CCCTTCTGTGGAACACCTTG
91	X_Contig121778_F	NW_007727769.1	TGCTGTCACATGTTGCCTTT	AAGTCTGGACCCCAGGTGTA
92	X_Contig24550_F	NW_007728142.1	GGAGTGGAGTCTGGTTCTGC	AATGCTGGGAACAGATGAGG
93	X_Contig13397_F	NW_007726755.1	GCGAATCGATTTCAAAAAGG	TCCCTAACACGGATTTCGAC
94	X_Contig5783_F	NW_007726442.1	CAGCTCTGTGAAGCAAGTGG	GGTTGAAGCATGTGGTGAGA
95	X_Contig15986_F	NW_007734964.1	ATTCTTCCCAAGAGCGACAA	TTCTCACAGCAAGAGCGTTG
96	X_Contig12283_F	NW_007726167.1	GACCCCATTCTGGAAACAAA	AGAGCAGAAGAGCAGGCTAGA
97	X_Contig77086_F	NW_007730981.1	CATGTCCAGATTCCTGCTCA	CCTCAACAGCAGGGATCAAT
98	X_Contig20651_F	NW_007726846.1	GATTTGGGCTGCCTGAGATA	TAGGCCAAATTGGGTCATCT
99	X_Contig33549_F	NW_007727546.1	CATGCCCCACACAGAACTTT	TGTTCACAGGGAAAGGGATT

Table C.6 Continued.

100	X_Contig11731_F	NW_007726833.1	GCGATAAGAATGCCCAAGAA	TGGACCCATGTTTTGAATCC
101	X_Contig609_F	NW_007727733.1	TTCCCTCAGAGTGAGCCTTT	GTGAAACCCCATGCACATTT
102	X_Contig46556_F	NW_007726521.1	TCGATAGAGTGCCCCTTTTG	CCCAGGTACTCGAATGAGGA
103	X_Contig42550_F	NW_007732640.1	TCTCGGGGAAATAGGGACTT	GACACTGGGGTGGGTTTCTA
104	X_Contig2858_F	NW_007729378.1	CCAGGCAACTGGGAGAATAA	CAGCAGCTCTCCCCTACCTA
105	X_Contig60171_F	NW_007727788.1	CGTAGCTGTTGCTTCCATCA	TCTCAGACCTTGTTGGCTCA
106	X_Contig7142_F	NW_007727922.1	ACCCACCTCCCCAACTATTC	TGCAATACTTTGGCTGCAAG
107	X_Contig27012_F	NW_007727409.1	AGACCTCCCCCATTCCTATG	TCCTTTGCTGCCCATTAAGT
108	X_Contig6097_F	NW_007726492.1	AGGTTTGCCAAAACCCCTAT	TGGCAGGAATTTCAGGCTAT
109	X_Contig16525_F	NW_007732233.1	CCAAAGAGCCCTTTAGTCCA	CCTTTTCATCATGCTGTGCT
110	X_Contig71242_F	NW_007726705.1	TCTTCTCTAGGCCCAACCAA	GCAGAACAAAAGCAGTTCCA
111	X_Contig19810_F	NW_007726193.1	GCCAAGAGACAGTGCCACAT	TAGCCTGAAGCCATCACCTT
112	X_Contig36839_F	NW_007730271.1	CCAATTACAAAGGGCTGCTC	GTTGCCTTGTGTGGCTGATA
113	X_Contig49850_F	NW_007726492.1	ATTTGCAAACCATTGGGAGA	TCTTCTCACTGACAGCCCAAT
114	X_Contig913_F	NW_007729102.1	AAGGTGTCACTCTGGGCTGT	CATTGGCTCTTGAGGCTTGT
115	X_Contig15702_F	NW_007728266.1	CAGAGGATCAGAGGGGTGAG	CTCTATGGAACCACCCAGGA
116	X_Contig10476_F	NW_007730510.1	CAACCTGATAAGGGACTTTTGC	TTCTGAAGTTCCCCATTGGA
117	X_Contig1173_F	NW_007726473.1	CACTGTGCAGATGCATTGAGT	CCCACTTTTGCACTGTTTGA
118	X_Contig261939_F	NW_007729855.1	TTAGCATCAACCCACACCAA	TTATTCGGCCACATTTCACC
119	X_Contig12541_F	NW_007732237.1	CCCTGTGTCCTACCATCACC	ATATCCCAGAGGGGGAAGAA
120	X_Contig4415_F	NW_007727377.1	CATGTGGAACTGCTTGACCA	AGGCAAGGCAAGAACAAGAA
121	X_Contig27403_F	NW_007726145.1	GAGTGAAGTCCCATGCTGGT	CCACTCGAACTGAGGGATGT
122	X_Contig36139_F	NW_007732597.1	GCTTCTTGGAACCTCTTTGG	TGCAAAAGATTCTGAGGCAGT
123	Y_Contig704_F	NW_007731050.1	GGCAGTTTGCTCAACTCTGG	TGGGAAAGTCATCGACAAAA
124	Y_Contig704_F	NW_007731050.1	TGCACAGTTCTCCACCTGAT	GGAACATTTTCCAACAGTTTTCA

Table C.6 Continued.

125	Y_Contig704_F	NW_007731050.1	CCTCCCAGGATCAAAAATCA	ATCTCAAATCATCCCCATCG
126	Y_Contig704_F	NW_007731050.1	GCTTGTCTGGGATCGTCTTT	TTTGAGGACATGAGGGTTGG
127	Y_Contig704_F	NW_007731050.1	AACATCAAGTGGGTGTTGTGA	GCCTTAATGAAATGATTTAAATTTCTG
128	Y_Contig704_F	NW_007731050.1	TCCTCTTTCCCTTCCTCCAT	CACAAATGGGTTGCTGATCT
129	Y_Contig75757_F	NW_007731050.1	CATGCATGAGAGCAAACACA	TCATTGCACTTGTCCCAGAG
130	Y_Contig1284_F	NW_007727559.1	TTGTGCACCTCCAAATTTCA	GCATAGAGGGAAGGCAGTTG
131	Y_Contig1284_F	NW_007727559.1	ATTCTGGTCCCACATGGAAG	GGTTTCTTTCTGCCTTGTCA
132	Y_Contig1284_F	NW_007727559.1	GGTGGTGATAAAAGGGTTTGG	CAAAACTACCCCCGGATTTT
133	Y_Contig1284_F	NW_007727559.1	TCTGTCTGGACTTCCTGTACCA	GTGGAGGGTCAATCCTGAGA
134	Y_Contig1284_F	NW_007727559.1	CAGGGAAAGACATGAATCGAA	CAAGCTGGTATCACCACTCCT
135	Y_Contig1284_F	NW_007727559.1	TGTGCAGAAACAAGAGGGACT	TCCAAGAGTTTGCAAGAAGGA
136	Y_Contig1284_F	NW_007727559.1	CGTGCGTGTGTAGAGAGAGAG	ATGAACAGCTGGAAGGGAAA
137	Y_Contig1284_F	NW_007727559.1	TCGCTCTTTTTCCTGATCAAA	TCTTCTCTCTCCCTCACCTCA
138	Y_Contig1284_F	NW_007727559.1	CTTGATTGCAGTGGGGTTTT	TTGCTTGTGACCAACTTTGTG
139	Y_Contig1284_F	NW_007727559.1	ATGAGCAAAGGTCAGCAACC	AGTCTAGGAACCTGGGGACAG
140	Y_Contig3524_F	NW_007733006.1	GTGGCCTTAACACCGTGAAT	CACCAGGACATGCAAGAAGA
141	Y_Contig3524_F	NW_007733006.1	CCAGAGGGAGAGGGTTTTCT	CCGATGGGGGAGTCTAATCT
142	Y_Contig3524_F	NW_007733006.1	CCTCACAGAAGGAGGTGTGG	GGGCTTGTCAACAGTTCTCC
143	Y_Contig3524_F	NW_007733006.1	CACCCCCACACACTTTTCTC	AGCCTTGGAATCATGAACAGA
144	Y_Contig3524_F	NW_007733006.1	AATTTTTCCATCAGGGAGCA	GGAGGTTAACAGCCCAGACA
145	Y_Contig3524_F	NW_007733006.1	CAGTTCTTCACGCCATCTGT	ATACCCCATCCAGACCCCTA
146	Y_Contig3524_F	NW_007733006.1	TGGTGTATGATCTTTCCTGTTGA	TGGCCAAGATAACACAATGG
147	Y_Contig15273_F	NW_007727559.1	TACACAAGCTCAAATACGTGACA	TTGCACATCAAGACATTTAGCA
148	Y_Contig29962_F	NW_007727559.1	CAGCATGCAAAATTGTTCTGA	ACGGGCTTGATTTTTAGCAC

Table C3.7. Morphometric measurements used for morphological principal component analyses.

Museum	sample number	Local	Dwarf	Sex	GSL	CBL	ONL	zyg	min w temps	P4-M3	bbc	hbc	P4W	M1W	M2W	M3W	max palw	diam orbit (d-V)	palate across T2s	p4-m3	mand condyle w	t1longL	t2longL
ZRC	4.993	PM	y	?	63.6	59.3	21.7	38.4	8.8	12	24.2	15.5	3.9	4	4.3	4.5	10.25	17.36	20	14.25	9.65	5.6	6.1
BMNH	94.8.2.1	THAI	n	?	68	63.8	23.5	42.55	9.1	14	24.7	18.2	4.6	5.1	5.1	5.1	11.2	18.8	21.7	15.7	9.4	6.7	6.6
BMNH	94.9.28.35	BNG	n	?	70	66.3	26.4	44	7	12.9	26.4	16.7	4.35	4	4.46	4.8	11.4	18	21.5	14.5	9.9	5.1	5.65
BMNH	46.332 (3375)	Unk	n	?	74.4	71.7	27.6	50.7	8.1	14.76	29.2	18.4	4.4	4.5	4.7	4.85	11.8	20.7	24.1	16.3	10.56	6.46	7
BMNH	50.11.22.33	PHI	n	?	66.1	63	22.1	44.8	2.1	15.7	24.6	18.5	5.66	5.6	6	5.8	11	16.8	23.2	16.9	9.8	5.3	5.9
USNM	112428	PM	y	f	66.8	62.5	24.3	42	8.7	13.1	26.7	16.5	4.2	4.1	4.6	4.75	10.8	17.6	20.8	14.3	9.6	5	6
ZRC	4.867	PM	y	f	69	65	24.4	42.4	8.6	13.3	25.3	16.9	3.7	4.1	4.3	4.66	11.4	18.1	22.2	14.6	10.5	4.8	6
BMNH	55.1447	PM	y	f	64.6	60.5	22.3	40.7	8.9	12.7	25.3	16.4	4	4.1	4.3	4.6	10.9	17.4	20.15	13.6	9.7	5.9	6.25
BMNH	55.1446	PM	y	f	65.8	62.2	23.2	41.2	9.65	12.9	25.5	16.1	4.1	4.2	4.7	4.6	10.7	17.8	20.4	14.3	9.7	5.9	6.2
ZRC	4.936	PM	y	f	65	62.2	22.8	42.6	9	12.9	24.5	16.7	4.1	4.6	4.8	4.95	11	19.1	21.25	14.2	9.46	5.8	6
BMNH	55.1439	PM	y	f	66.4	63.3	23.8	43.9	9.1	12.8	26.7	17.35	4.4	4.8	4.85	4.83	11.3	19.3	21.3	15	9.2	5.4	5.5
ZRC	4.931	PM	y	f	60.88	58.6	21.4	39.1	8.37	12.2	23.4	17	3.4	3.9	4.17	4.14	10.3	18.3	19.5	13.7	9	4.9	5.35
ZRC	4.932	PM	y	f	64.1	61.5	22.6	42.9	9.8	12.8	25.3	16.8	4.25	4.18	4.69	4.8	10.8	18.7	20	15.5	10.5	5	5.7
ZRC	4.933	PM	y	f	65.9	62.5	23.6	43.3	10.2	12.3	23.5	16.35	4.28	4.23	4.9	4.5	11.7	18.5	20	14.1	10	4.3	5.25
BMNH	55.1438	PM	y	f	64.4	62.2	22.7	42.4	8.6	13	24.8	16.6	4.3	4.2	4.4	4.3	11.3	18.4	21.1	14.2	9.3	5.1	5.6
ZRC	4.958	PM	y	f	63.1	59.8	22.6	41.9	12	13	24.7	17.4	4	4	4.3	4.6	10.3	18.2	19.9	14.6	9.4	5.1	5.35
ZRC	4.960.	PM	y	f	68.6	64.1	24.5	41.6	13.4	12.5	25.1	16.7	3.9	4.1	4.6	4.7	10.9	18.8	20.2	14.4	9.4	5.3	6
BMNH	47.1427	PM	y	f	66.1	62.8	24.5	43.3	13	12.3	26.1	17.2	4	4	4.8	4.6	10.5	17.7	21	13.8	8.9	5.6	5.4
BMNH	9.11.1.13	PM	y	f	65.2	61.9	23.4	43.8	10.44	12.3	27	17.4	3.9	3.7	4.3	4.4	12	18.3	21.3	13.9	10	5.6	5.6
ZRC	4.953	PM	y	f	62.2	58.9	22.1	40.8	12.8	12.4	25.4	17.2	3.9	4	4.3	4.3	10.3	18.34	20.3	14.1	9.08	5.5	5.2
BMNH	47.1428	PM	y	f	66.1	62.5	23	42.8	7.7	12.8	23.2	17	3.7	3.8	4.4	4.2	10.8	18.2	19.7	14.74	8.4	5.4	6.2

Table C3.7 Continued.

BMNH	**9.4.1.98**	**SUM**	**n**	**f**	**71.1**	**66.2**	**27**	**43.9**	**7.8**	**13.1**	**26.2**	**17.6**	**4.1**	**4.5**	**4.9**	**4.6**	**11.1**	**17.9**	**21.55**	**15**	**8.7**	**6**	**6.6**
BMNH	9.4.1.97	SUM	n	f	72.3	67.7	26.9	45.6	10.7	13.1	26.7	18.1	4.1	4	4.5	4.5	11.8	18.2	22.1	15.2	9.4	5.1	5.7
USNM	125124	W.BOR	y	f	56.5	53.9	20.9	33.3	10.6	11.9	22.8	16.3	3.13	3.85	4.07	4.26	8.9	16	17.6	13.4	7.4	5	4.75
ZRC	4.979	SUM	n	f	75.7	72.3	29.3	47.6	10	15.1	28.1	17.1	4.8	4.8	5	5.3	11.9	20.2	23.8	17.2	10.9	6.4	6.4
ZRC	4.868	SUM	n	f	68.3	64.5	25.7	42.4	9.5	12.76	25.4	18.2	4	4	4.6	4.9	11.5	19.4	21.1	14.1	8.8	5.7	5.8
ZRC	4.870.	BNG	n	f	69.8	66	23	43.5	9.8	13	26.3	17.6	3.9	4.1	4.6	4.7	11.6	19.3	22.1	14.66	8.8	5.5	5.8
ZRC	4.871	BNG	n	f	73	67.7	27.3	47.9	10.2	13.6	28.8	17.9	4.2	4.5	4.6	4.8	11.6	19	22.4	14.9	9.15	5.8	6.3
ZRC	4.872	BNG	n	f	70.3	67.4	26.8	45.6	9.05	13	28	18.5	3.9	4.1	4.3	4.8	10.5	19.5	22	14.5	10.5	6.2	6.2
USNM	104601	NAT	n	f	66.9	64.2	26.5	39.4	7.5	13.1	23.5	14.8	4.1	3.9	4.5	4.5	11.3	16.5	20	14.4	8.7	5.9	6.3
ZRC	4.875	NAT	n	f	67.2	64.8	25.3	42	7.7	12.8	25.4	15.8	4.1	4	4.2	4.2	10.8	18.5	21	15.9	9.9	6	6.3
ZRC	4.877	NAT	n	f	63.7	60.5	24.7	40.6	7.7	12.3	25.1	16.5	3.8	3.9	4.1	4.1	10.4	17.9	20.2	14.1	8.3	5.75	6.1
ZRC	4.878	NAT	n	f	68.9	66.1	26.7	42.9	6.5	13.4	24.1	16.5	4	4	4.3	4.4	11.8	18.1	21	15	9.7	6.2	6.1
ZRC	4.985	NAT	n	f	65.8	62.5	25	43.3	4.3	13	22.9	16.8	4.5	4.4	4.65	4.6	11.8	18.1	22.6	14.66	9.6	6	6.4
ZRC	4.882	ANAMB	n	f	68.2	63.2	24.2	43	9.9	12.9	26.3	16.76	4.37	4.61	4.93	4.36	10.8	19.2	22.3	14.65	8.6	5.4	5.76
ZRC	4.884	ANAMB	n	f	72.9	69.5	27.3	44.3	8.9	13.5	25.9	17	4.5	4.77	5.35	4.94	10	19.4	21.6	15.4	9.5	5.93	6.7
USNM	145577	W.BOR	n	f	63.3	60.45	22.4	40.4	9.8	12.1	23.2	15.65	3.5	3.75	4.17	3.96	10.76	17.7	20	13.6	8.5	5.55	5.78
AMNH	107137	W.BOR	n	f	68.2	65.5	24.5	42.7	9.35	13.2	25.7	17	4.3	4.3	4.5	4.8	11.9	18.6	20.2	14.3	9.2	6.75	6.6
ZRC	4.893	NE.BOR	n	f	70.6	66.3	23.9	48	11.8	14.37	26.3	18.4	4.2	4.67	5	5	11.4	20.1	21.5	16.8	8.9	5.6	5.9
ZRC	4.894	NE.BOR	n	f	74.2	70.1	27.3	47.5	12.6	14.2	27	18.34	4.06	4.38	4.77	4.93	12.7	19.2	22.8	15.95	9.55	7	6.76
USNM	197202	E.SE.BOR	n	f	74.65	70.3	26.4	44.6	13.4	13.3	27.6	17.2	4.2	4.4	4.6	4.6	11.2	19.45	22.1	15	8.4	6.7	6.7
USNM	198050	W.BOR	n	f	73.8	69	29.3	46	11.9	14	28.8	18.5	4.4	4.6	5	5.1	11.8	19.6	24.3	15.9	9.4	6.3	6.55
USNM	151888	NE.BOR	n	f	74.1	69.6	26.4	46.3	13.1	14.1	27.8	17.8	4.3	4.6	5.1	4.9	13.5	19.6	23.9	15.7	8.9	6	6.4
AMNH	103735	NE.BOR	n	f	72.2	70.1	29.1	48.7	9.9	13.3	28.2	18.3	4.3	4.3	4.6	4.6	12.5	21	22.5	15.3	10.2	5.5	6.6
BMNH	55.1449	PM	n	f	73.8	68.8	28.7	45.1	7.7	13.8	26.6	17.9	4.7	4.7	5.35	5.1	12	17.8	23.7	15.85	10.2	7.05	7.2

Table C3.7 Continued.

BMNH	**55.1450.**	**PM**	**n**	**f**	**74.3**	**69.9**	**28.75**	**45.9**	**10.3**	**14**	**28.1**	**18.4**	**4.5**	**4.9**	**4.9**	**5.1**	**12.9**	**19.1**	**23.3**	**15**	**10.1**	**7.1**	**7.35**
ZRC	4.939	PM	n	f	73.6	69.7	28.4	44.7	10.46	14.8	26.8	17.8	4.7	4.9	5.4	5.1	11.5	18.56	22.6	16.3	9.9	7.2	7.2
ZRC	4.940.	PM	n	f	71.55	67.6	28.4	43.8	9.1	13.6	25.6	18.5	4.5	4.3	4.9	4.9	11.9	18	22.6	15.1	9.2	6.05	6.5
ZRC	4.938	PM	n	f	77.35	71.8	28.2	46.9	11.05	14.8	27.5	18.05	5.12	4.72	5.23	5	11.9	19.35	23.1	17.8	10.4	7.7	7.7
BMNH	8.7.20.10, type of peninsulae	PM	n	f	74.8	71.5	27	49.9	11.9	15.2	27.1	19.1	4.5	4.8	5.1	5.15	11.9	21.1	24.2	16.2	10.1	6.3	5.9
ZRC	4.902	PM	n	f	72.1	67.9	28.5	43.6	9.3	14.04	27.9	17.5	4.7	4.7	5.06	5.2	11.7	20	22.7	16	9.8	6.2	6.4
ZRC	4.909	PM	n	f	76.75	73.6	28.6	46.3	11.2	13.2	28.6	18.1	4.1	4.2	4.6	4.6	12.35	19.7	24.3	15.5	10.2	6.3	6.3
ZRC	4.905	PM	n	f	76.4	73.6	27.7	48.2	11.6	14.1	29.5	19.2	4.74	4.56	5	5.13	12.78	21.4	25.7	16	9.8	7	7.5
ZRC	4.907	PM	n	f	75.7	73	28.1	49.1	10.9	13.66	28.3	18.3	4.3	4.5	4.8	4.8	13.2	20.9	24.8	15.2	10.05	6.08	6.2
ZRC	4.908	PM	n	f	75.25	72.7	28.8	50.2	9.5	13.4	27.2	20.3	4.6	4.65	4.66	5.08	13.87	20.2	24.4	15.6	10.2	5.76	5.75
USNM	307552	PM	n	f	74.4	70.9	28.7	44.7	9.5	13.8	28.5	19.3	4.65	4.1	4.6	4.7	12.6	18.4	24.3	14.9	10.2	7	7.15
ZRC	4.920.	PM	n	f	78.5	74.1	30	52	11.7	13.9	31.7	19.5	4.3	4.3	4.7	4.9	12.6	20.4	23.8	15.6	10.6	5.7	6.1
USNM	84920	THAI	n	f	76.7	72.2	29.2	49.9	13.1	14.6	27.8	18.8	4.6	5	5.2	5.2	13.5	21	13.5	16	11.4	6.1	6.2
USNM	255716	THAI	n	f	73.6	69.7	25.4	47.9	8.6	13.5	26.3	18.3	4.5	4.7	4.9	4.9	12.65	20.9	23	15.3	10.6	6	6.6
AMNH	85139	THAI	n	f	73	70.4	27.8	46.6	8.1	14.6	27.6	18	4.4	4.7	4.7	5.4	12.3	18.9	23	16.2	10.5	6.6	6.5
ZRC	4.944	SUM	n	f	74.9	71.5	28.6	47.2	8.9	14.5	26	18.55	5.16	4.75	5.21	5.38	12.2	20	24.7	16.8	10.2	6.8	7
ZRC	4.947	SUM	n	f	73.75	71.8	26.7	46.9	9.5	13.3	27.6	17.4	4.46	4.3	4.55	4.5	12.8	20.5	21.6	15.1	9.06	5	5.9
BMNH	7.6.18.3	SUM	n	f	70	66.5	24.5	42.2	7.9	13.8	24.3	17.5	4.4	4.1	4.6	4.5	11.1	18.5	21.4	14.7	9.5	5.8	5.5
BMNH	7.6.18.2	SUM	n	f	69.4	66.1	23.4	44.2	4.4	13.6	25.5	17.9	4.2	4	4.3	4.36	10.9	18.8	20.9	15.3	9.9	5.3	6.1
ZRC	4.962	JAV	n	f	79.3	76.3	27.6	50.3	13	15.6	27.4	19.5	4.9	5.3	5.65	5.4	14	21	25.5	17.1	10.6	7.5	7.9
AMNH	101501	JAV	n	f	72.6	69.7	28	45.3	11.5	15.2	25.6	18	4.8	4.9	5.3	5.4	10.6	20.4	22.1	16.5	9.9	7	6.7
USNM	112427	PM	y	m	60.2	57.3	21.6	38.6	10.55	13.1	25.1	17.2	4	4.4	4.7	4.75	9.6	17.1	18.7	14.1	8.4	5.3	6.2
BMNH	55.1445	PM	y	m	61.2	57.9	20.8	40.2	9.3	12.5	26.1	16	3.9	3.9	4.45	4.6	9.75	18	19.3	13.9	8.7	5.3	5.9

Table C3.7 Continued.

BMNH	**8.1.25.26**	**PM**	**y**	**m**	**65.9**	**62.9**	**23.9**	**40.1**	**8.9**	**12.9**	**25.7**	**16.2**	**3.9**	**4.1**	**4.4**	**4.4**	**10.2**	**16.9**	**20.3**	**14.5**	**9.4**	**5.6**	**6.1**
ZRC	4.866	PM	y	m	62.1	58.7	20.5	40.5	6.95	13	24.7	16.3	3.9	4.1	4.45	4.7	10	17.9	20.4	14.3	9.1	5.2	5.2
ZRC	4.934	PM	y	m	60.7	56.6	20.7	42.5	9.3	12.5	24.1	17.7	4.27	4	4.37	4.16	10.2	18.7	18.9	13.2	9.1	4.2	5.2
ZRC	4.935	PM	y	m	60.8	58.5	22	39.6	10.2	12	22.9	16.7	3.9	4.16	4.6	4.5	9.57	18.2	19.1	13.5	8.4	5.3	5.5
ZRC	4.957	PM	y	m	62.7	59.2	23.55	41.4	11.3	12.2	23.75	17.36	3.8	3.8	4.2	4.3	10.3	18.46	19.8	13.9	8.6	4.9	5.6
ZRC	4.951	PM	y	m	60.4	56.1	21.7	40.6	11.7	12.9	23.9	16.4	3.9	4.1	4.6	4.56	10.4	17.9	19.3	14.5	8.5	5.2	5.5
BMNH	55.1440.	PM	y	m	57.8	54	20.4	39.4	11.4	12.2	27	16.7	3.75	4.1	4.45	4.66	9.3	17.9	19	13.5	8.2	4.8	5.3
BMNH	55.1441	PM	y	m	60.7	57.7	22.1	38.1	11.2	13	24.2	16.4	4	4.2	4.4	4.5	9.8	17	19.7	14.1	8	5.5	5.5
BMNH	55.1442	PM	y	m	64.8	61.8	24.3	43.8	10.7	12.4	26	16.9	3.9	4	4.5	4.85	10.85	19	19.9	13.6	8.8	4.4	5.55
ZRC	4.959	PM	y	m	64.7	60.9	22.5	41.5	9.6	13.8	22.6	16.5	3.7	4.4	4.7	4.3	10.5	19.4	19.1	14.5	8.7	5.1	5.5
ZRC	4.961	PM	y	m	61.9	57.9	21.4	40.6	10.9	12.4	24	17	4.2	4.1	4	4.7	10.7	18.2	19.7	13.1	8.3	5.8	6
BMNH	9.11.1.11	PM	y	m	64.7	60.6	23	41.3	8.9	12.9	26.3	17.2	4.2	4.2	4.3	4.7	11.4	18.3	19	14.9	8.8	5.8	6.2
BMNH	9.11.1.10	PM	y	m	62.5	58	24.1	38.5	9.8	11.7	23.4	16.1	3.3	4	4.2	4.1	9.8	17	18.5	13.9	8.7	5.5	5.5
ZRC	4.927	PM	y	m	62.1	59	21.2	40.3	8.6	12	23.2	16.6	4.2	4.2	4.5	4.5	9.7	17.4	19.55	14.05	8.7	5.4	6.1
ZRC	4.929	PM	y	m	63	59.3	21.5	40.8	8.75	12.9	20.3	15.6	4.5	4.2	4.5	4.65	9.6	17.7	19.3	14.2	9.2	5.5	6
ZRC		PM	y	m	62.1	59	22.2	41.9	10.6	12.8	23.1	16.2	4.2	4.3	4.6	4.6	10.4	18	20.2	14.6	9	5.2	6.2
USNM	123035	SUM	y	m	63.2	59.8	23.3	38.5	7.8	12.3	24.6	15.2	3.9	4.3	4.5	4.3	9.8	17.7	19.65	13.45	8.5	6.3	6
USNM	123069	SUM	n	m	66.3	63	25.1	41.9	8.8	13.1	26.5	17.1	4.1	3.9	4.35	4.5	11	17.5	19.8	15	8.2	5.4	5.6
BMNH	9.4.1.96	SUM	n	m	66.3	62.8	25	43.4	8.3	13	24.9	18	3.8	3.8	4	4.5	10.2	18.4	20.2	14.2	8.9	5.7	6.1
BMNH	9.4.1.95	SUM	n	m	63.5	59.8	23.5	40.5	8.9	13.4	25.8	16.9	4.2	4	4.1	4.3	11	17.6	20.3	14.1	7.9	5.8	6.2
FMNH	171074	LAOS	n	m	63.1	59.6	19.6	42.6	10.8	12.25	21.8	16	3.7	4	4.4	3.3	10.3	19.5	19.1	13.9	8.3	5	5.9
USNM	114376	SUM	y	m	65	62	24.3	39.4	12.4	13.1	26.1	16.4	3.9	4.1	4.4	4.2	11.1	18.2	19.9	14	7.9	6	6
ZRC	4.981	SUM	n	m	72.4	68.2	26.4	47.7	8.2	14.05	27.4	17.8	4.4	4.1	4.8	5.1	11	20.3	22.1	15.3	10.2	6.2	6.7
ZRC	4.874	BNG	n	m	66.8	63.2	25.7	43.2	8.3	12	25.7	15.6	3.9	4.1	4.3	4.3	11	19	21.8	13.9	9.5	5.75	6

Table C3.7 Continued.

USNM	**104600**	**NAT**	**n**	**m**	**62.9**	**59.8**	**22.7**	**39**	**10.2**	**11.8**	**23.7**	**16.6**	**4**	**3.9**	**4**	**4.25**	**10.1**	**17.9**	**20**	**13.4**	**8.7**	**5.9**	**6**
ZRC	4.880.	NAT	n	m	61.2	58.1	20.8	41.7	8.5	12.3	22.4	16.8	3.9	4.2	4.5	4.4	11.3	18.3	20.8	13.95	8.3	6.5	6.1
USNM	151887	E.SE.BOR	n	m	65.43	61.05	24.4	41.1	8.9	13.2	23.5	17.9	4	4.3	4.6	4.65	9.7	18.1	19.45	14.9	8.1	6.05	6.2
ZRC	4.896	NE.BOR	n	m	66.5	63.2	24.3	42.2	12.1	13.7	23	17.8	3.65	4.36	4.59	4.59	11	18.9	19.5	15.3	8.6	5.4	6.1
ZRC	4.895	NE.BOR	n	m	68.3	65.55	24.9	43.6	10.8	14.3	25.4	17.3	4.27	4.17	4.5	4.7	10.9	18.05	19.8	15.3	8.8	6.3	6.55
ZRC	4.891	NE.BOR	n	m	69.8	66.8	25.9	45.1	11.7	13.66	26.3	16.8	3.96	4.1	4.44	4.44	10.77	19.35	21.9	15.3	9	5.9	6.1
ZRC	4.888	NE.BOR	n	m	74.4	69.5	25.8	48.1	12.9	13.6	27.8	18.4	4.1	4.3	4.8	4.7	13.2	19.7	23.7	15.6	9.4	5.8	6.2
ZRC	4.866	NE.BOR	n	m	68.9	66.2	24.2	44.1	10.9	12.2	24.2	17.9	4.07	4.23	4.25	4.44	11.3	19.8	21.1	14.9	8.63	6.12	5.84
BMNH	0.8.4.4	NE.BOR	n	m	68.5	64.7	25.7	43.3	13.4	12.7	26.2	17.5	3.8	4.16	4.3	4.3	11.8	19.6	20.5	14.3	8.1	5.6	5.6
BMNH	0.2.2.6	NE.BOR	n	m	64.8	60.7	22.6	42.8	9.4	12.5	26.6	18.1	3.7	4.2	4.85	4.6	10.6	19.6	20.5	13.6	7.9	5.45	6.25
USNM	198051	E.SE.BOR	n	m	68	63.8	24.5	45.6	14	13.6	28.2	18.75	4.7	4.7	5.2	5.4	11.5	19.5	21.2	14.75	9.3	6.15	6.4
USNM	176431	E.SE.BOR	n	m	70.2	66.4	26.2	43.6	14.3	13.6	25.5	17.1	4.3	4.4	4.7	4.8	11	19	21.6	15.9	9.3	5.9	6.4
USNM	198704	E.SE.BOR	n	m	69.6	64.9	26	43.1	14.1	12	25	17.2	3.5	3.8	4	4.1	11.8	18.4	20.8	14	8.55	5.55	5.9
AMNH	103734	NE.BOR	n	m	72.4	67.8	24.9	46.2	15.6	11.5	27.9	18.4	4.1	4.1	4.6	4.5	11.3	20.7	21.2	14.2	9.7	5.8	6
AMNH	106286	W.BOR	n	m	59.7	56.9	20.9	39.5	12.8	11.7	24.8	18	3.5	3.7	3.9	4.1	9.7	18.4	18.7	13.1	7.5	5.4	5.6
ZRC	4.941	PM	n	m	69.9	67.3	25.7	42.2	8.4	13.7	24.85	17.2	4.5	4.4	4.4	4.56	11.15	18.4	21.2	14.9	10.5	5.7	6.27
ZRC	"425"	PM	n	m	69.75	66.2	24.6	44.5	7.1	15	23.56	19.1	4.74	5.16	5.55	5.28	11.2	20.6	23.2	15.86	8.3	7.16	7.14
ZRC	4.937	PM	n	m	66.8	63.4	25.8	42.4	11.6	14	23.3	17.8	4.3	4.2	4.56	4.74	10.8	18	20.8	14.9	8.7	6.24	7.2
BMNH	55.1448	PM	n	m	72	69	27.8	45.2	7.4	14.3	25.9	18.8	4.7	4.7	5.2	5.1	12	19	20.8	16	10	5.4	5.9
ZRC	4.906	PM	n	m	72.1	69.2	27	45.4	11.7	12.9	26.4	18.1	4.2	4.6	4.76	4.68	11.5	19.7	23.7	15.2	9.75	5.7	6.1
ZRC	4.899	THAI	n	m	71.2	67.1	27	45.5	11.1	13.35	28.25	19.6	4.4	4.3	4.9	5.3	11.9	20.3	21.2	15.3	9.5	6.2	6.5
ZRC	4.925	PM	n	m	71.8	68.8	27.4	46.9	10.8	14.4	28.1	17.5	4.3	4.7	5.05	4.7	10.76	19.9	21.7	15.7	9.9	5.7	6.5
ZRC	"446"	PM	n	m	72.5	70.1	27.9	45.4	10.3	14.2	29.1	18.4	4.5	4.6	5.15	5.25	11.5	19.7	22.2	16	10.4	6.3	6.7
AMNH	106629	SUM	n	m	74.7	70.9	26.4	47.2	11.7	15	24.4	18.3	4.4	4.8	5	5.1	13.1	19.4	23.6	16.7	10.6	6.7	6.1

Table C3.7 Continued.

AMNH	**106630**	**SUM**	**n**	**M**	**74.2**	**71**	**28.7**	**48.7**	**10.7**	**14.3**	**26.1**	**18.1**	**4.9**	**5**	**5.1**	**5.1**	**10.7**	**20.6**	**23.6**	**16.5**	**10.1**	**6.5**	**6.9**
USNM	121749	SUM	n	m	77.1	75	27.9	49.7	8.8	15.65	29.1	19.4	4.7	5.1	5.6	5.6	13.3	21.4	24	17.4	11.1	5.6	7.05
USNM	121747	SUM	n	m	72.8	70.3	27.4	49.4	13	15.7	26.1	17.8	5	4.9	5.4	5.3	13.04	21.3	23.7	17	10.15	5.8	6.8
BMNH	7.6.18.4	SUM	n	m	66.8	66.4	24.2	44.1	7.4	12.5	24.6	18.3	4.2	3.9	4.2	4.4	8	19.2	20.6	13.8	9.95	6.1	5.8
ZRC	4.965	JAV	n	m	81.3	76.4	31.8	50	10	15	27.9	18.6	5.17	5.1	5.76	5.47	14.1	20.9	24.8	17.8	11.8	6.05	7.2
ZRC	4.968	JAV	n	m	74.6	70.7	29.7	46.3	12.1	14.9	26.8	18.2	4.7	5.4	5.4	5	11.8	20.4	22.9	15.8	8.8	6.2	7.4
BMNH	1938.11.30.73	JAV	n	m	81.9	76.5	29.7	50.4	11.3	15.2	29.9	18.9	5.5	5.3	6.3	5.75	13	21.4	24.6	17.5	10.7	7.2	7.8
AMNH	101500	JAV	n	m	69.9	67.1	27.1	45.5	9.9	14.3	24.5	18.1	4.5	4.6	5	4.7	10.4	19.9	22.2	15.2	9.2	5.8	6.2
BMNH	8.1.25.25	PM	y	m	60.3	57.6	22.4	38.4	9	13.4	25	16.4	4.3	4.5	4.8	4.8	9.26	17.1	18.9	14.8	8.8	5	6