1-1-2013

Platyrrhine Phylogenetics With A Focus On Callitrichine Life History Adaptations

Natalie Mae Jameson
Wayne State University,

PLATYRRHINE PHYLOGENTICS WITH A FOCUS ON CALLITRICHINE LIFE HISTORY ADAPTATIONS

by

NATALIE MAE JAMESON

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2013

MAJOR: MOLECULAR BIOLOGY & GENETICS

Approved by:

_____

Advisor Date

_____

_____

_____

_____

DEDICATION

To my soon-to-be husband Chris, who knew that most bad days could be remedied with a crab leg dinner.

ACKNOWLEDGMENTS

Firstly, I thank my adviser, Derek Wildman. The education I received and the opportunities were afforded to me over my time in the Wildman lab will no doubt drive me to success throughout my career.

Secondly, thank you to my committee members, Dr. Heng, Dr. Uddin, and Dr. Yi. It is because of their continual input and support that this research has been shaped into a body of work I am deeply proud of. Thank you to the staff of the Center for Molecular Medicine and Genetics, without whose help and backing I would have been lost. I will be forever grateful to the Wildman lab members, both past and present. Their valuable advice and unwavering friendships have been vital in carrying me through this process. I honor the memory of Dr. Morris Goodman, a brilliant scientist and wonderful man. I will always feel privileged for the time that I was able to learn from him.

Finally I'd like to thank my family and friends. You know who you are and you know all you've done. Thank you, I love you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Understanding reproductive strategies is a central aspect in studies of species evolution (Kappeler et al., 2003). The unique assemblage of a species' life history traits defines its reproductive strategy. Life history traits having a major influence on reproductive strategy are those such as, number versus quality of offspring, current versus future reproduction, age versus size at maturity, and fecundity versus lifespan. Drastic changes in traits such as these occur on a long evolutionary time scale (i.e. between Orders), while smaller adjustments to these traits can occur in shorter time (i.e. within an Order) (Promislow and Harvey, 1990). However there are exceptions, such as litter size within the Order *Primates*. Within *Primates*, litter size varies from 1 to 4 offspring per pregnancy; with the majority of species producing singleton offspring (Leutenegger, 1979). Interestingly, species within *Cebidae* (squirrel monkey, capuchin monkey, owl monkey, marmoset, tamarin, and Goeldi's monkey) produce litters that span the full spectrum of one to four offspring (Ross, 1991). The variation in litter size seen among the relatively closely related *Cebidae* species suggests an alteration of reproductive strategies within the past 10 million years (Opazo et al., 2006; Perelman et al., 2011; Jameson et al., In Review, See chapter 4).

The aim of my research is to integrate phenotypic and genotypic character evolution in order to determine the molecular underpinnings of reproductive strategies in New World monkeys. I have employed the use of comparative genomics, tracing the

evolution of genes and phenotypes on the background of a well resolved phylogeny to determine conserved and derived character states, which has been shown to be a valuable method to study both phenotypic and genotypic evolution. However the precision of this approach is directly dependent on the accuracy of the phylogenetic framework. To this end, I have taken the following steps in building the necessary phylogenetic frameworks and identifying the molecular underpinnings of reproductive strategies in callitrichines.

In the first aim (Chapter 2) we set the basis for the phylogenetic framework. Using a phylogenomic approach we generated multiple sequence alignments of 1,078 orthologous genes among 17 mammal species, from which we determined the optimal phylogenetic tree topology and estimated divergence dates. The goal of this study was to determine the closest relative to anthropoid primates. The results provided overwhelming support for the presence of a haplorrhine clade, a grouping that includes anthropoid primate and tarsiers to the exclusion of strepsirrhine primates. With a properly rooted anthropoid phylogeny we then set out to determine the phylogenetic and biogeographic history of New World monkeys (NWM). As previous studies have described (Peng et al., 2009; Wildman et al., 2009), non-genic sequences are powerful phylogenetic markers, therefore, in aim 2 we created a bank of genomic sequence data from which to extract non-genic sequences (Chapter 3). We generated genomic shotgun libraries from three NWM species (Spider monkey, Owl monkey and Uakari), which produced 3,154 individual sequence reads. We also provided detailed annotation and human-marmoset alignments of all individual and assembled sequences. The availability of this genomic data allowed us to continue to aim 3 by providing a source of

non-genic sequence data (Chapter 4). Here our goal was to develop a well supported NWM phylogeny upon which we could estimate divergence times and trace the biogeographic history of the taxa. We generated a non-genic DNA multiple sequence alignment of 40,986 base pairs including indels among 36 NWM species. Phylogenetic analysis of this data matrix produced a tree topology with 88% of the nodes fully supported. We were also able to determine the geographic area of the most recent common ancestor of extant NWMs, and using divergence time estimates, examine the paleogeography and paleoclimate during that time period. In addition, our phylogenetic analysis confirmed Goeldi's monkey (*Callimico goeldii*) as sister taxa to the marmosets (*Callithrix*) nestled within the callitrichines. These results form the foundation for the last aim, identifying the molecular underpinnings of varying reproductive strategies (Chapter 5). To do this we implemented a comparative genomics approach and sequenced the transcriptomes of four NWM species (Goeldi's monkey, *Callimico goeldii*; lemurine owl monkey, *Aotus lemurinus*; coppery titi monkey, *Callicebus cupreus*; and black-headed spider monkey, *Ateles fusciceps*) using next-generation sequencing technologies. We tested for signs of adaptive evolution in 97 genes known to be involved in reproduction through previous studies. Through these analyses we detected genes undergoing adaptive evolution coincident with the changes in reproductive strategy, on the terminal branches of Goeldi's monkey and marmoset as well as on their stem branch. To test if dual ovulation, a trait present in all other callitrichines, was maintained in Goeldi's monkey we looked for evidence of chimerism in a study population. Two independent techniques were both unable to detect non-self alleles, suggesting that the mechanism of litter size reduction occurs pre ovulation.

The following section of my thesis will serve as background into my specific aims. I will provide support from the literature for the work that has been done and give an introduction to the concepts discussed in the work. In addition each chapter contains its own topic specific background and introduction.

New World monkey phylogenetics

Primates are thought to have reached the neotropics via a rafting event that carried them across the Atlantic Ocean from Africa to South America sometime during the Oligocene (Ciochon and Chiarelli, 1980a; Houle, 1999). These primates underwent a period of rapid divergence during which they adapted to the new environments, filling the novel niches now available to them. This primate clade contains over 125 extant New World monkey species (Goodman et al., 2005; Groves, 2001, 2005; von Roosmalen and von Rossmalen, 2003). The phylogenetics of New World monkeys has been studied for decades, with early studies focusing on the morphological and ecological differences among the species (Ford, 1986; Kay, 1990; Rosenberger, 1981, 1984). The results of these studies were widely varied, and thus provided limited phylogenetic resolution for the clade.

Subsequent phylogenetic analysis of NWMs focused on molecular datasets using sequence from both the nuclear and mitochondrial genomes; however, results from these studies were rarely in agreement (Barroso et al., 1997; Canavez et al., 1999; Chaves et al., 1999; Goodman et al., 1998; Harada et al., 1995; Horovitz and Meyer, 1995; Opazo et al., 2006; Porter et al., 1997a, 1999; Schneider et al., 2001; von Dornum and Ruvolo, 1999). Inability of these methods to converge on a single topology may be largely due to the use of datasets of limited size and restricting the datasets to

sequences that have been under strong selection pressures, such as coding and mitochondrial DNA. The subsequent use of *Alu* insertions as a marker for phylogenetic change by Ray et al. (2005) and Osterholz et al. (2009) provided more resolution; however, *Alu* elements have low phylogenetic power due to their inability as markers to infer dates of divergence or rates of nucleotide substitution.

Non-genic DNA markers

Here we have used non-genic markers developed from non-coding, non-repeat, intergenic genomic DNA sequence as a tool for phylogenetic analysis. Previous studies suggest that these markers appear to be a more powerful tool than other techniques that have previously been used for phylogenetic reconstruction (Wildman et al., 2009). This power is a result of their neutrally evolving nature that keeps them free from the constraints of Darwinian selection. Conserved noncoding mammalian sequences have been found to evolve at a rate 1.7 times faster than nonsynonymous sites, and undergo lineage specific relaxation (Nikolaev et al., 2007). This provides a more accurate phylogenetic picture by tracking neutral/random lineage specific molecular changes among the species rather than molecular changes due to adaptation to the surrounding environments.

Reproduction strategies within the Cebidae

Given that the ability to pass genetic information from generation to generation is the basis of fitness among organisms, many reproductive strategies in which to accomplish this have evolved and been optimized over time. Within the *Cebidae* family of New World monkeys there are two distinct reproductive strategies in terms of litter size. Members of the *Cebinae* (*Cebus* and *Saimiri*) along with *Aotus* produce singleton

offspring. *Aotus* and *Saimiri* produce a single offspring every year, while *Cebus* typically has a two-year inter-birth interval (Dixon, 1983; Fragaszy and Adams-Curtis, 1998; Rosenbloom, 1968). *Cebus* and *Aotus* offspring are occasionally transported by their fathers a few weeks into life; however *Saimiri* offspring are only carried by the mothers (Garber and Leigh, 1997; Valenzuela, 1994; Wright, 1981). In contrast, most genera within *Callitrichinae* (*Callithrix*, *Leontopithecus* and *Saguinus*) have increased their reproductive output by producing litters of 2-4 offspring as often as twice per year (Goldized, 1987; Tardif et al., 2003). These 43 New World monkey species are the only anthropoid primates to consistently give birth to dizygotic, chimeric twins, triplets, or quadruplets. Excluded from this group is *Callimico*, the only callitrichine that gives birth to singleton offspring (Altmann et al., 1988).

Twinning and Chimerism

Early evidence for hematopoetic chimerism in marmosets (*Callithrix jacchus*) and tamarins (*Saguinus fusicollis* and *Leontopithecus rosalia*) was described by Benirshcke et al. (1962), Gengozian et al. (1964), and Niblack et al. (1977), and recently germ line chimerism has been suggested to occur in *Callithrix kuhlii* (Ross et al., 2007). In the 45 years between these studies a variety of evidence has been presented linking callitrichine primates and inter-sibling chimerism. Marmosets are natural blood chimeras; a condition resulting from a high frequency of fraternal twinning, chorionic fusion and the consistent development of placental vascular anastomoses allowing exchange of fetal blood between the two embryos. In the best-studied callitrichine species, *Callithrix jacchus* (the common marmoset), twin blastocysts implant at day 12 of gestation, and chorionic fusion of the two early placentas takes place between days

19 and 29, after which blood and genetic material can be exchanged among the developing fetuses (Enders and Lopata, 1999; Moore et al., 1985). In addition to callitrichines, cattle have also shown the ability to produce chimeric twin offspring. This condition known as the freemartin syndrome arises when vascular connections form between the placentas of developing heterosexual twin fetuses and XX/XY chimerism develops resulting in varying degrees of masculinization of the females' tubular reproductive tract and negative effects such as irremediable sterility (Padula, 2005). This outcome is dissimilar to the effects of chimerism on marmosets, which when chimeric, display immunological benefits including specific immune tolerance to co- twin antigens (Benirschke et al., 1962; Niblack et al., 1977; Wislocki, 1939). Macrochimerism in humans is produced by errors in the fertilization process such as the fusion of two different zygotes in a single embryo and results in two genetically distinct cell lines in an individual. Microchimerism, the presence of 1-5% allogeneic donor cells, is often a byproduct of organ transplantation, pregnancy or transfusion. Although no positive impact has been directly linked to maternal fetal chimerism, long-term transplant tolerance has been associated with the presence of hematopoietic microchimerism (Starzl et al., 1993). Chimerism can eliminate the need for life long immunosuppressive drugs post-transplant by ensuring intrathymic T cell deletion of donor-reactive cells for as long as the chimerism persists (Pree et al., 2007). Naturally occurring germ cell chimerism has been detected in the colonial urochordate, *Botyrllus schlosseri* (Pancer et al., 1995), and in the cnidarian, *Hydracitina symbiolongicarpus*, although the latter species does not have circulating blood and there is no evidence of tolerance induction by embryonic chimerism across histocompatibility barriers (Poudyal et al., 2007).

Single born callitrichines

Marmosets and tamarins are very energetically conservative and can alter their litter size based on their perceived availability of resources. In captivity, where resources are plentiful, marmosets give birth to litters of triplets and quadruplets; in the wild, where resources are often more scarce, twins are most common as well as the occasional birth of a chimeric singleton (Rutherford and Tardif, 2008). The resorption of a twin resulting in a singleton birth is seen in 23.8% of marmoset pregnancies and is thought to be a direct consequence of depleted resources resulting in poor maternal nutrition (Jaquish et al., 1996). The finding of chimerism in single born marmosets demonstrates the resorption of one fetus late in gestation after it has contributed hematopoetic stem cells to its twin (Gengozian and Batson, 1975). The resorption of late term embryos during the second half of gestation without ill effect to the mother has only been described in marmosets. Although this phenomenon is rare, twin embryo resorption is also seen in cattle, when a freemartin singleton is born following an ultrasound verifying twin pregnancy. In humans this phenomenon is known as the vanishing twin effect, and is seen in 21% of 1000 human pregnancies during the first trimester (Landy et al., 1986).

***Callimico* reproduction**

In contrast to marmosets and tamarins, *Callimico goeldii* is the only callitrichine that gives birth to a single offspring (Altmann et al., 1988). This feature along with the presence of a third molar is largely responsible for the debate surrounding the phylogenetic position of *Callimico*. In contrast to morphological studies that place *Callimico* as the basal callitrichine (Ford, 1986; Kay, 1990; Kay, 1994; Rosenberger and

Coimbrafilho, 1984; Snowdon, 1993), genetic based studies have been able to place *Callimico* well within the callitrichines as the sister taxa to *Callithrix* (Canavez et al., 1996; Canavez et al., 1999; Chaves et al., 1999; Cronin and Sarich, 1978; Horovitz and Meyer, 1995; Opazo et al., 2006; Pastorini et al., 1998; Schneider et al., 1993; Seuanez et al., 1988; von Dornum and Ruvolo, 1999; Wildman et al., 2009). Based on the topology derived from genetic studies, the most parsimonious explanation for singleton births in *Callimico* is that twinning evolved once on the stem of callitrichines and was subsequently lost on the *Callimico* terminal lineage after divergence from *Callithrix* ten million years ago (Opazo et al., 2006). We hypothesize that due to environmental stresses incurred on the *Callimico* terminal lineage the reduction of litter size possibly via embryo resorption became advantageous to the survival of the offspring, resulting in the birth of chimeric singletons.

The development of and care for a second fetus through much of gestation by *Callimico* would be energetically costly for the mother. Moreover, while marmoset fathers carry one twin and mothers carry the other for the first few months after birth, *Callimico* doesn't produce twins and paternal carrying of the infant is delayed up to 4 weeks after birth (Carroll, 1982; Jurke et al., 1994; Pook, 1977). This distinction would result in an increase in the *Callimico* mother's energy expenditure during the perinatal period. This energetically costly practice is potentially a trade-off outweighed by the benefits of singleton births in the genus. Because *Callimico* gives birth to singletons, the mother can exclusively carry the infant for the first weeks of its life. The delay in alloparental care of *Callimico* offspring is due to maternal prevention, not the unwillingness of the fathers, as alloparental care decreases the infant's safety by

increasing the risk of predation and accidents (Schradin and Anzenberger, 2001). *Callimico* infants are also at less risk of predation because they are born larger and grow at a more rapid rate than *Callithrix* infants. Larger body size and rapid growth in *Callimico* may be achieved because the infant is the sole recipient of nutrients and care by the mother (Ross et al., 2010). In the case of chimeric marmosets and tamarins the amount of relatedness between parents, offspring, and siblings is increased beyond that typically observed because it is possible for offspring to contain genetic information from all paternal and maternal alleles (Haig, 1999). The increase in relatedness between the family members could promote the increased investment in offspring by the father and siblings. This increase in alloparental care may be of great benefit for their survival into adulthood and reproductive maturity. Because chimerism can help explain the evolution of alloparental care in callitrichines, it is important to discover whether the alloparental care exhibited by *Callimico* is due to continued chimerism or is merely an evolutionary hangover from the past twinning in the evolutionary history of the genus.

In Summary

Many factors can affect the evolutionary history of a species, possibly none more so than species reproductive strategy. The following chapters will discuss the evolutionary history of anthropoid primates and will attempt to connect phenotypic variations to their genotypic origins. I will describe two independent phylogenomic approaches; both successful in producing well supported primate phylogenetic frameworks. In addition, the geographic ranges of NWMs have been mapped onto the phylogeny, thus elucidating the biogeographic history of the clade. I will also present an investigation into alterations of reproductive strategy among callitrichines. Using a

comparative genomics approach, I identify genes adaptively evolving on the lineages where twinning evolved, and was subsequently lost. Finally, through the detection of chimerism, I consider the possibility of embryo resorption as the mechanism for the loss of twinning.

CHAPTER 2

Genomic data reject the hypothesis of a prosimian primate clade

This chapter has been published as:
Jameson, N.M., Hou, Z.C., Sterner, K.N., Weckle, A., Goodman, M., Steiper, M.E., Wildman, D.E., 2011. Genomic data reject the hypothesis of a prosimian primate clade. Journal of human evolution 61, 295-305.

Summary

The phylogenetic position of tarsiers within the primates has been a controversial subject for over a century. Despite numerous morphological and molecular studies there has been weak support for grouping tarsiers with either strepsirrhine primates in a prosimian clade or with anthropoids in a haplorrhine clade. Here, we take advantage of the recently released whole genome assembly of the Philippine tarsier, *Tarsius syrichta*, in order to infer the phylogenetic relationship of *Tarsius* within the order Primates. We also present estimates of divergence times within the primates. Using a 1.26 million base pair multiple sequence alignment derived from 1,078 orthologous genes, we provide overwhelming statistical support for the presence of a haplorrhine clade. We also present divergence date estimates using local relaxed molecular clock methods. The estimated time of the most recent common ancestor of extant Primates ranged from 64.9 to 72.6 Ma, and haplorrhines were estimated to have a most recent common ancestor between 58.9 and 68.6 Ma. Examination of rates of nucleotide substitution in the three major extant primate clades show that anthropoids have a slower substitution

rate than either strepsirrhines or tarsiers. Our results provide the framework on which primate morphological, reproductive, and genomic features can be reconstructed in the broader context of mammalian phylogeny.

Introduction

Tarsius (Tarsiiformes: Primates) is the most distinct extant primate genus. Tarsiiformes (i.e., tarsiers) include extant members of the genus *Tarsius*, as well as their extinct relatives, and are estimated to have diverged from all other living primates in the late Paleocene epoch (Goodman, 1999). Moreover, the unique combination of morphological traits in tarsiers has impeded their placement within the primate phylogenetic tree (Yoder, 2003). Despite uncertainty regarding the phylogenetic affinities of *Tarsius*, most researchers consider that there are three main extant primate clades: 1) Anthropoidea (platyrrhines and catarrhines), 2) Strepsirrhini (Malagasy lemuriforms and lorisiforms), and 3) Tarsiformii (*Tarsius*). Assuming monophyly of these three groups leads to four primary hypotheses regarding the phylogenetic relationships among them (Figure 2.1). The first hypothesis posits a prosimian clade and/or grade (Prosimii) in which *Tarsius* is grouped as the sister taxon to the Strepsirrhini (Chatterjee et al., 2009; Hasegawa et al., 1990; Murphy et al., 2001). Conversely, the second hypothesis places *Tarsius* as sister to the Anthropoidea and together these primates form the clade Haplorrhini (Baba et al., 1982; Goodman et al., 1998; Pocock, 1918; Schmitz et al., 2001; Zietkiewicz et al., 1999). The third hypothesis places *Tarsius* as sister taxon to a clade comprised of Strepsirrhini and Anthropoidea (Andrews et al., 1998; McNiff and Allard, 1998). Finally, the fourth hypothesis suggests that *Tarsius*,

Anthropoidea, and Strepsirrhini diverged rapidly from one another effectively resulting in a trichotomy at the base of Primates (Arnason et al., 2008).

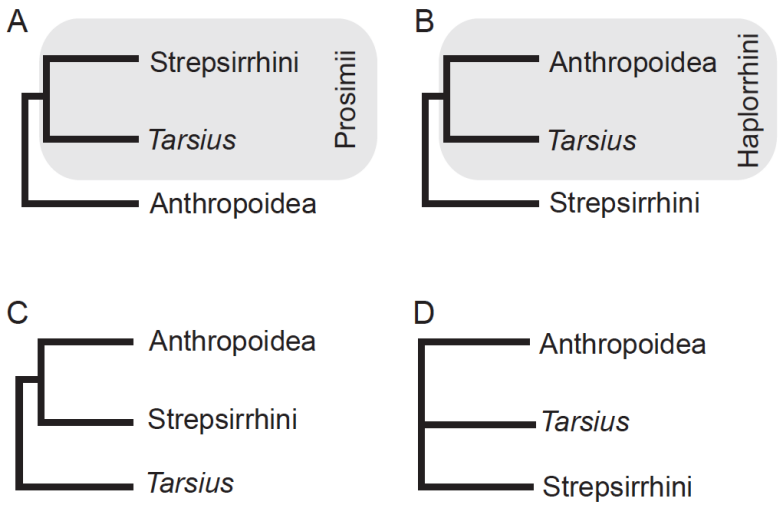


Figure 2.1: Four hypotheses for the phylogenetic relationships among the three major primate clades. (A) Tarsius groups with Strepsirrhini to form the prosimian clade, (B) Tarsius groups with Anthropoidea to form the haplorrhine clade, (C) Anthropoidea groups with Strepsirrhini to the exclusion of Tarsius, and (D) a trichotomy among the three major primate clades.

Understanding the phylogenetic relationships among tarsiers, anthropoids, and strepsirrhines has proved challenging due to weak or conflicting phylogenetic signals in morphological and molecular datasets. This lack of resolution prevents accurate reconstructions of the evolution of the phenotypes and underlying genotypes that distinguish these lineages. Clearly, a well-supported, well-sampled primate phylogeny is needed to address questions fundamental to our understanding of primate evolution. For example, reconstructing the morphotype and ancestral genome sequence of the last common ancestor of extant anthropoids requires knowledge of the closest

anthropoid outgroup. Until this gap in knowledge is filled, reconstructing the anthropoid ancestor remains a challenge.

Tarsiers have a deep evolutionary history, with fossil evidence suggesting origins extending to the base of the modern primate radiation. While the data are relatively sparse, including mainly isolated teeth and mandibular fragments, paleontological records show the presence of several fossil tarsiers in Asia that date to the Eocene and Miocene epochs (approximately 34-56 Ma and 23-5 Ma respectively) (Beard, 1998b; Beard et al., 1994; Gradstein et al., 2004; Mein and Ginsburg, 1997). Three species have been identified from these fossils: *Tarsius thailandicus* (Thailand; (Mein and Ginsburg, 1997)), *Tarsius eocaenus* (Jiangsu Provence, China; Beard et al., 1994), and *Xanthorhysis tabrumi* (Shanxi Provence, China; Beard, 1998b). There is also evidence that at least one extinct tarsier-like species may have lived in Africa. The Egyptian fossil *Afrotarsius chatrathi* is estimated to have lived approximately 32 Ma during the Oligocene and was identified as a tarsiid primate based on craniodental data (i.e., five teeth and a single mandible; Simons and Brown, 1985) as well as a controversial piece of postcranial data (i.e.. tibiofibula; Rasmussen et al., 1998). These fossils establish the existence and antiquity of the tarsier clade, but do not resolve the branching order among the three major primate clades. Several independent lines of evidence have been interrogated in an attempt to reconstruct the evolutionary relationships of the tarsiers, anthropoids, and strepsirrhines. The major findings inferred from both morphological and genetic data are summarized below.

Morphological characters among Tarsius*, Anthropoidea and Strepsirrhini*

Often tarsier features have been considered to represent a primitive primate state, and thus tarsiers have been grouped with strepsirrhines in the “lower primate” grade (Simpson, 1945). This classification is supported by features shared between tarsiers and strepsirrhines such as presence of a grooming claw, an unfused mandible symphysis, small body size, saltatory locomotion, and nocturnal lifestyle (Niemitz, 1983; Schwartz, 2003). However, tarsiers also share features with anthropoids, including a dry rhinarium, a fused lateral cleft of their nasal processes, and reduced olfactory bulbs (Pocock, 1918; Rosenberger and Szalay, 1980). Although tarsiers are nocturnal (similar to many strepsirrhines), their eye anatomy is more comparable to those of anthropoids in the number of photoreceptors, absence of a tapetum lucidum, and the structure of the postorbital septum (Hendrickson et al., 2000; Martin, 1973; Pocock, 1918). In terms of reproduction, tarsiers are similar to anthropoids in that they have a hemochorial placenta, similar sperm morphology, and display delayed puberty (Groves, 1986; Luckett, 1974; Pocock, 1918; Robson et al., 1997; Wislocki, 1929). However, tarsiers are the only primate to have both a hemochorial placenta as well as a bicornuate uterus (Schwartz, 2003). With an average gestation length around 178 days (Izard et al., 1985) and an average adult body mass of 130 g (Roberts, 1994), tarsiers have the longest gestation length relative to body size of any living mammal genus. Morphological data appear to carry conflicting phylogenetic information regarding the placement of *Tarsius* relative to Anthropoidea and Strepsirrhini. The character state distribution for select features among the three extant primate clades is summarized in Table 2.1 without reference to character state polarity. Tracing the evolution of these character states

requires a well-inferred phylogenetic tree that includes a wide range of suitable non-primate outgroup taxa (Maddison et al., 1984).

Table 2.1
Presence (X) of select character states that vary among the three major primate clades.

		Extant			
	Feature	Tarsiers	Strepsirrhines	Anthropoids	Reference
Skeletal	Presence of grooming claw	X	X		Beard 1988
	Elongated tarsal	X	X		Beard 1988
	Tibiofibular syndesmosis	X	X		Beard 1988
Jaw/Teeth	Absence of toothcomb	X		X	Hill 1953
	Unfused mandible	X	X		Jablonski & Crompron 1994
Auditory	Ontogenetic internalization of the ectotympanic	X	X		Cartmill and Kay 1978
	Primary supply of the middle meningeal artery via maxillary artery	X		X	Cartmill and Kay 1978
Metabolic	Unable to synthesize vitamin C	X		X	Pollock and Mullin 1987
Nasal	Dry rhinarium	X		X	Pocock 1918; Rosenberger & Szalay 1980
	Fused lateral nasal cleft	X		X	Pocock 1918; Rosenberger & Szalay 1980
	Reduced olfactory bulbs	X		X	Pocock 1918; Rosenberger & Szalay 1980
	Rostral elongation	X	X		Stark 1984
Vision	Orbital enlargement with medial constriction	X	X		Pocock 1918; Martin 1973; Hendrickson 2000
	Increased number of photoreceptors	X		X	Pocock 1918; Martin 1973; Hendrickson 2000
	Absence of tapetum lucidum	X		X	Pocock 1918; Martin 1973; Hendrickson 2000
	Apical interorbital septum	X		X	Cave 1973
	Structure of postorbital septum	X		X	Pocock 1918; Martin 1973; Hendrickson 2000
	Fovea centralis	X		X	Wolin & Massoupust 1970
	Macula lutea	X		X	Polyak 1957
Reproduction	Bicornuate uterus	X	X		Luckett 1974; Mossman 1987
	Hemochorial placenta	X		X	Pocock 1918; Wislocki 1929; Luckett 1974; Groves 1986; Wildman 2006; Elliot 2009
	Small/absent allantois	X		X	Mossman 1987
	Delayed puberty	X		X	Pocock 1918; Groves 1986; Robson 1997

Genetic evidence for the phylogenetic placement of Tarsius

Genetic data have also been used for reconstructing phylogeny among primate clades. Phylogenetic analyses of both mitochondrial and nuclear sequence data have been used to infer the position of *Tarsius*. As with morphological data, previous molecular studies have not yielded a consensus among these relationships. A number of molecular studies have been used to support a prosimian clade. Two independent phylogenetic analyses found evidence of a prosimian clade within primates using the same mitochondrial (ND3/ND4 genes) dataset (Hasegawa et al., 1990; Hayasaka et al., 1988), although it has been suggested these results may have been biased due to improper taxon sampling (Yang and Yoder, 1999). Recent studies of complete mtDNA genomes have found limited evidence for a haplorrhine clade, but only when the unique

base composition of the tarsier mtDNA genome is accounted for in the analyses (Matsui et al., 2009). Phylogenetic analysis of nuclear DNA (using a region of the αA-crystallin gene) has also been used as evidence supporting a prosimian clade (Jaworski, 1995). This same study, however, also suggested prosimians are more closely related to bats, rodents, and treeshrews, resulting in a non-monophyletic primate clade. A more extensive dataset including 10 kb of nuclear and mitochondrial data also concluded that tarsiiforms were sister taxa to lorisiforms and lemuriforms, but again resulted in a non-monophyletic primate clade placing flying lemurs as the sister taxa to Anthropoidea (Murphy et al., 2001). Most recently, phylogenetic analyses of both a nuclear and a mitochondrial gene supermatrix suggested support for a prosimian clade (Chatterjee et al., 2009), although it should be noted that within the study *Tarsius* was only represented by the mitochondrial supermatrix. A similar study using a supermatrix combining both nuclear and mitochondrial genes placed tarsier outside of an anthropoid + strepsirrhine grouping; however this topology was not significantly different from topologies that included monophyletic Prosimii or Haplorrhini clades (Fabre et al., 2009). In some instances mitochondrial data have also been used to support a non-primate positioning of tarsiers (Andrews et al., 1998; McNiff and Allard, 1998).

While some molecular evidence supports a prosimian clade, molecular data have also provided evidence for a monophyletic haplorrhine clade. To date, research based on the β-globin gene cluster represents the vast majority of data that support a Haplorrhini-Strepsirrhini division (Baba et al., 1982; Bailey et al., 1991; Goodman et al., 1998; Koop et al., 1989; Meireles et al., 2003; Page and Goodman, 2001; Porter et al., 1995). In addition to the globin studies, other genomic regions provide support for a

haplorrhine clade (Bonner et al., 1980; de Jong and Goodman, 1988; Djian and Green, 1991). Moreover, insertions of Alu transposable elements have been used in primate phylogenetic reconstruction. Evolutionary relationships inferred from Alu data (based on the presence or absence of shared insertions) show three Alu markers to be derived features uniting tarsiers and anthropoids, thus supporting the division of primates into Haplorrhini and Strepsirrhini (Schmitz et al., 2001; Zietkiewicz et al., 1999).

While genetic data can be powerful for phylogenetic analysis, in many cases these data have been unable to conclusively resolve the branching order among the three primate groups. Moreover, the rapid rate at which mtDNA evolves, compared to the slower rate of nuclear DNA evolution, can confound efforts to infer relatively ancient species divergences, due to factors such as long branch attraction and substitution saturation (Felsenstein, 2004). As larger datasets are able to provide more phylogenetic signals, a large, slowly evolving nuclear dataset would be valuable in accurately resolving the position of *Tarsius* in relation to Anthropoidea and Strepsirrhini. Indeed Simpson (Simpson, 1945) predicted that complete genetic data would be “priceless” in inferring phylogeny. The recent explosion of genome sequencing in humans and other mammals thus provides the raw data necessary to solve difficult phylogenetic problems. Several recent phylogenomic studies (Decker et al., 2009; Hallstrom et al., 2007; Nishihara et al., 2007; Prasad et al., 2008; Wildman et al., 2007) have illuminated the relationships among the major mammal clades, and these approaches could be applied within Primates.

The recent public release of the draft assembly of a low-fold coverage genome from the Philippine tarsier (http://genome.wustl.edu/genomes/view/tarsius_syrichta/),

Tarsius syrichta, provides an opportunity to test primate phylogenetic relationships using phylogenomic methods. Here, we present the results of maximum parsimony (MP), maximum likelihood (ML), and Bayesian analyses performed on two datasets. Each dataset consisted of orthologous transcripts from 17 mammalian species including anthropoids, tarsier, strepsirrhines, and outgroups. In the larger of the two datasets, each species was represented by 1,078 concatenated transcripts (up to 1.26 million base pairs). The second dataset employed a more restricted, extensively curated multiple sequence alignment containing 47 concatenated transcripts per species (45,000 base pairs). From the results, we infer the phylogenetic branching pattern and the divergence dates among the primate lineages represented by our study taxa. We then discuss how our findings may help us better understand the course of primate evolution.

Material and methods

Genome sequences and alignment quality control

Sequences were obtained for 17 mammalian taxa (Table 2.2) from the Ensembl comparative database (v. 56, October 2009). We used the Ensembl comparative protein homology database, which is limited to only the longest transcripts for each gene and relies solely on Ensembl predicted gene datasets. In order to avoid complications from inclusion of gene duplication and conversion events, we used only 1:1 (single copy) orthologs as determined by Ensembl (Vilella et al., 2009). This very conservative approach examines only those genes with relatively simple evolutionary histories. All gene families with multiple copies in a given genome were excluded from our analyses.

Table 2.2

Mammalian taxa used to infer the phylogeny

Order	Common Name	Scientific Name	Fold Coverage	Genome Source
Primates	Human	*Homo sapiens*	Complete	Genome Reference Consortium
	Common Chimpanzee	*Pan troglodytes*	6	GSC and the Broad Institute
	Orangutan	*Pongo pygmaeus*	6	Washington University Genome Sequencing Center
	Rhesus monkey	*Macaca mulatta*	5.2	Macaque Genome Sequencing Consortium, Baylor College
	Common Marmoset	*Callithrix jacchus*	6	Washington University Genome Sequencing Center
	Tarsier	*Tarsius syrichta*	1.82	Washington University Genome Sequencing Center
	Mouse lemur	*Microcebus murinus*	1.93	Broad Institute
	Galago	*Otolemur garnetti*	1.5	Broad Institute
	Tree shrew	*Tupaia belangeri*	2	Broad Institute
Lagamorphs	Rabbit	*Oryctolagus cuniculu*	2	Broad Institute
	Pika	*Ochotona princeps*	1.93	Broad Institute
Rodentia	Mouse	*Mus musculus*	Complete	Mouse Genome Sequencing Consortium
	Rat	*Rattus norvegicus*	11.9	Baylor College
	Kangaroo rat	*Dipodomys ordii*	1.85	Human Genome Sequencing Center Baylor College & Broad Institute
	Guinea pig	*Cavia porcellus*	6.79	Broad Institute
Cetartiodactyla	Cow	*Bos taurus*	7	Human Genome Sequencing Center Baylor College
Carnivora	Dog	*Canis familiaris*	7.6	Broad Institute

We generated two datasets that varied in sequence quality. First, we created a dataset of 1,087 orthologous transcripts. A second smaller dataset was then created by removing sequences with more than 10% undetermined nucleotide bases ('Ns') and those with a total length less than 150 bp in any of the 17 species. This smaller and more rigorously curated dataset included 47 orthologous transcripts. For both datasets, cDNA alignments were generated using predicted protein sequences using Bioperl modules (http://www.bioperl.org) within MUSCLE (Edgar, 2004). Because most of the genomes used in this study have less than 2X coverage, both datasets included some level of low quality sequence that might significantly reduce the alignment quality and affect phylogenetic inferences (Talavera and Castresana, 2007). In order to account for this, we chose to eliminate poorly aligned positions and excessively divergent genomic regions from the phylogenetic analyses, as these regions are often not homologous or have been saturated by multiple substitutions (Castresana, 2000; Talavera and Castresana, 2007). All sites removed were done so at the cDNA level and included the whole codon in order to maintain the proper reading frame. cDNA sequences were curated using Gblocks-v0.91 with option –t=c (codon) and all other parameters set to

default such that the minimum number of positions in agreement for a conserved position was set to 50% +1, the minimum number of sequences for a flank position was set to 85%, the maximum number of contiguous non-conserved positions was set to 8, the minimum length of a block was set to 10, and no gap positions were allowed (Castresana, 2000). High quality alignment regions were retained and concatenated using a custom perl script. Finally, for the larger dataset, 1,078 of the 1,087 transcripts met the minimum conserved sequence criteria (conserved cDNA sequence length is longer than 50 base pairs) and were retained for analysis. All 47 transcripts in the smaller dataset met the minimum conserved sequence criteria, and the higher sequence quality filter criteria and were retained for analysis.

Phylogenetic analyses

Maximum parsimony (MP) analyses were conducted on both the 47 transcript and the 1,078 transcript datasets in PAUP* (Swofford, 2002). Gaps in the alignment were treated as missing data. One thousand bootstrap replicates were performed with 10 random addition sequence heuristic replicates using the TBR branch-swapping algorithm.

Maximum likelihood (ML) analyses were conducted on both datasets in the parallelized MPI-enabled version of RAxML-7.23 (Stamatakis, 2006). The best-known likelihood tree was inferred after conducting 100 RAxML runs with a random starting tree. One thousand bootstrap replicates were performed, and nucleotide substitution models were chosen, as detailed below. MrAIC.pl (Nylander, 2004) was used to select the best nucleotide substitution model for each individual gene according to the Akaike Information Criterion (AIC) (Figure S1). The 47 transcript dataset was partitioned using

individually selected models; the 1,078 transcript dataset was run as a single partition using the model most frequently chosen among the transcripts (GTR+Γ, 73%). To test whether alternative tree topologies differed from each other statistically, Shimodaira-Hasegawa (SH) topology tests (Shimodaira and Hasegawa, 1999) were conducted in PAUP*. Each dataset was run with the GTR+Γ +I model as chosen by MrAIC.pl (Nylander, 2004).

Bayesian inference (BI) was implemented in MrBayes-3.1.2, recompiled as an MPI version (Ronquist and Huelsenbeck, 2003) for 500,000 MCMC generations with a burn-in of 2,500 generations as determined by convergence statistics. Each analysis consisted of two runs, with four chains and sample tree. Chains were sampled per 1000 generations. Nucleotide substitution models were implemented as in the ML analysis (i.e., the 47 transcript dataset was partitioned and the 1,078 transcript dataset was run as a single partition using GTR+Γ). In all phylogenetic analyses, dog and cow were used as outgroups to root the tree.

Divergence time estimation

Divergence dates were estimated from the 1,078 transcript dataset described above, as well as from two previously published datasets described below. A relaxed molecular clock was applied to the three partitions. The first partition comprised only third codon positions from the 1,078 transcript dataset and included 422,687 bp. Third positions were chosen because they are less likely to be constrained by selection relative to the pressures faced by first and second codon positions. The second partition was derived from a 1.9 million base pair region of human chromosome 7 surrounding the *CFTR* locus (Cooper et al., 2005), then reduced to a 59,764 bp region, as described

in Steiper and Young (Steiper and Young, 2006). In the present analysis, this partition included these 59,764 bp positions for 17 taxa (Table S1). The third partition is derived from the *CYP7A1* region, located on human chromosome 8, and has been used for dating divergence times among primate lineages (Wang et al., 2007; Wilkinson et al., 2011). This dataset consisted of orthologous sequence from 9 taxa (Table S1) and included 22,906 bps. The two additional partitions were added to increase the taxonomic coverage, allowing for more nodes to be dated. Cumulatively, 505,357 bp of data from 26 species were used for divergence date estimations with no overlap of transcripts among the three datasets.

These three data partitions were analyzed individually using the *mcmctree* program of *PAML* v.4.4c (Yang, 2007). The approximate likelihood method of *mcmctree* was used for the large dataset because this method is much faster than the full Bayesian method of *mcmctree*. On a dataset of this size, a fully Bayesian approach would have taken weeks to run. For each analysis, branch lengths were estimated using *baseml* using the HKY+Γ_5 model (Hasegawa et al., 1985; Yang, 1994) with different transition/transversion rate ratios (κ), base frequencies, and gamma shape parameter α values used for the three partitions.

For the approximate likelihood Bayesian molecular clock analysis, the following parameters were used. The substitution rates were assumed to drift over time independently at the three loci, i.e., a “relaxed” molecular clock. The time unit used was 100 Ma. The rate at the root was assigned the gamma prior $\mu \sim G(0.125, 0.5)$, with mean of 0.25, corresponding to 2.5×10^{-9} substitutions per site per year. The geometric Brownian motion model was used to accommodate the drift of the substitution rate over

time (across lineages) (Rannala and Yang, 2007). A model of correlated rate change among lineages was used. The rate-drift parameter was assigned the gamma prior $\sigma^2 \sim G(10, 100)$. The prior for times is generated from the birth-death process with sampling, with parameters λ (birth rate) = μ (death rate) = 2 and ρ (sampling fraction) = 0.1. The following tree was used: ((((((((((*Homo*, *Pan*), *Gorilla*), *Pongo*), (((*Macaca*, *Papio*), *Chlorocebus*), *Colobus*)), (((*Saimiri*, *Aotus*), *Callithrix*), *Callicebus*)), *Tarsius*), ((*Microcebus*, *Lemur*), *Otolemur*)), *Tupaia*), ((*Ochotona*, *Oryctolagus*), (((*Mus*, *Rattus*), *Dipodomys*), *Cavia*))), ((*Canis*, *Felis*), *Bos*)).

Two different methods of fossil calibration were used to estimate divergence times. In the first method referred to as our bounded fossil calibrations the oldest fossil attributable to a key clade was used as a minimum age for that particular clade. These dates are considered "hard" lower bounds, as there is little chance that the estimated dates could be more recent than these divergences. Using this method, the minimum age at the *Homo*/*Pan* divergence was calibrated as 7 Ma or older, based on the age of *Sahelanthropus* (Lebatard et al., 2008). The *Macaca*/*Papio* divergence was calibrated as 6 Ma or older, based on the age of the earliest members of *Macaca* (5.5 Ma) and paleontological estimates for the divergence of these taxa between 7 – 8 Ma (Delson, 1992; Delson et al., 2000). The crown Catarrhini node was calibrated as 20.6 Ma or older, based on the age of *Morotopithecus* (Gebo et al., 1997). The crown anthropoid divergence was calibrated as 34.5 Ma or older, based on the age of *Catopithecus* (Seiffert, 2006). All of these hard lower bounds used a "soft" truncated Cauchy distribution for an upper bound with parameters $p = 0.1$ and $c = 1$ (Inoue et al., 2010).

The second calibration method, referred to as our probabilistic calibrations, is based on the recent study of Wilkinson et al., (Wilkinson et al., 2011). This method uses data on the rates and times of primate fossil appearances from the fossil record to conduct a stochastic modeling approach to generate non-uniform distributions for divergence times of the crown primate and crown anthropoid nodes. This method generates a calibration distribution that has tails of limited probability at both the lower and upper ends of the distribution, along with estimates of higher probabilities at the middle of the distribution. Here, we used Wilkinson et al.'s (Wilkinson et al., 2011) Poisson sampling model. In this method, the calibration prior distribution corresponds to a mean divergence of 56 Ma for the anthropoid node and 71 Ma for the primate node. The lower 2.5% bound is at ~45 Ma for the anthropoid node and ~61 Ma for the primate node, and the upper 2.5% bound is at ~68 Ma for the anthropoid node and ~88 Ma for the primate node.

Outside of the primates, fossil calibrations used as hard lower bounds were 43 Ma for the *Canis*/*Felis* divergence (Wesley-Hunt and Flynn, 2005) and 53 Ma for the *Ochotona*/*Oryctolagus* divergence (Rose et al., 2008). These hard lower bounds used a soft upper bound from a truncated Cauchy distribution with parameters p = 0.1 and c = 1. Two other bounded calibrations were used with a hard lower bound, and an upper bound with a limited probability (2.5%) that the date estimate can be older, i.e., a "soft bound" (Yang and Rannala, 2006). The *Mus*/*Rattus* calibration bound was from 10.4 Ma to 14 Ma (Jacobs and Flynn, 2005). *Cavia* of the remaining rodents was calibrated as greater than 50 Ma, based the occurrence of chapattimyids (Hartenberger, 1982),

and less than 65 Ma (Asher et al., 2005; Ivy, 1990). The root node of the tree was given a maximum age of 125 Ma.

For both calibration models, two independent MCMC runs were initiated from different starting points using an approximate likelihood. Each chain began with a burnin of 500,000 steps, followed by 500,000 samples drawn every 100 steps, for a chain length of 50 million. The acceptance rates for each of the estimated parameters were tuned to between 0.2 – 0.45. Convergence between runs was assessed by examining the parameter estimates between the two runs, examining trace plots, and calculating the effective sample sizes. The results presented are the mean and 95% equal tail credibility interval from the posterior distributions of two independent runs, averaged.

In addition to divergence date estimation via mcmctree, we also conducted analyses in BEAST v. 1.6.1 (Drummond et al., 2012). For this analysis we estimated the divergence dates of 17 taxa using the 47 transcript dataset. This analysis was done using a relaxed uncorrelated lognormal molecular clock, with a Yule Process tree prior and calibrations equivalent to those used in mcmctree. The analysis was run for 20 million generations, and sampled every 1000 generations.

Results

Data composition

After implementing quality control measures the larger alignment consisted of 1,268,061 bp positions derived from 1,078 concatenated transcripts across 17 taxa. The smaller alignment consisted of 45,948 bp positions derived from 47 concatenated transcripts sampled across the same 17 taxa. Because bias in nucleotide composition may lead to incorrect tree inferences (Nei and Kumar, 2000), we examined the

nucleotide composition for each species (Table S2). In the high quality dataset, GC composition ranged from 50.6% (*Callithrix jacchus*) to 54% (*Ochotona princeps*). In the 1,078 transcript dataset, GC composition ranged from 46.5% (*Callithrix jacchus*) to 50.2% (*Ochotona princeps*). This degree of nucleotide composition difference should not influence the phylogenetic relationships inferred from these data.

Phylogenetic analyses

Trees inferred from both datasets using all three methods (parsimony, likelihood, and Bayesian) converged on a single topology within primates, with *Tarsius* as the sister taxa to Anthropoidea (Figure 2.2). These data support separation of the primates into the two suborders: Haplorrhini (tarsier, New World monkeys, Old World monkeys, and apes) and Strepsirrhini (lemuriforms and lorisiforms). At the node connecting *Tarsius* to Anthropoidea, maximum parsimony (MP) bootstrap percent scores for both datasets = 100, and maximum likelihood (ML) bootstrap percent score is 79 for the 47 transcript dataset and 100 for the 1,078 transcript dataset. Bayesian inference (BI) posterior probabilities at this node for both datasets = 1.0. Among all mammals examined, these trees are identical with one exception, the position of the treeshrew (*Tupaia belangeri*). Maximum parsimony analysis of the larger dataset suggests that the treeshrew is sister to the Glires (bootstrap = 99), while all other analyses place *Tupaia* as sister to the primates, (47 transcript dataset: MP/ML/BI = 61/100/1.0; 1,078 transcript dataset: ML/BI = 79/1.0, see Figure 2.2). Both datasets support Eurochontoglires, Glires, and Primates as monophyletic clades, while the 47 transcript dataset also supports the monophyly of Euarchonta. After MP analysis, the 1,078 transcript dataset

was found to have a significantly lesser proportion of parsimony informative sites (24%) than in the 47 transcript dataset (32%) (p<0.001, two-tailed Fisher's exact test).

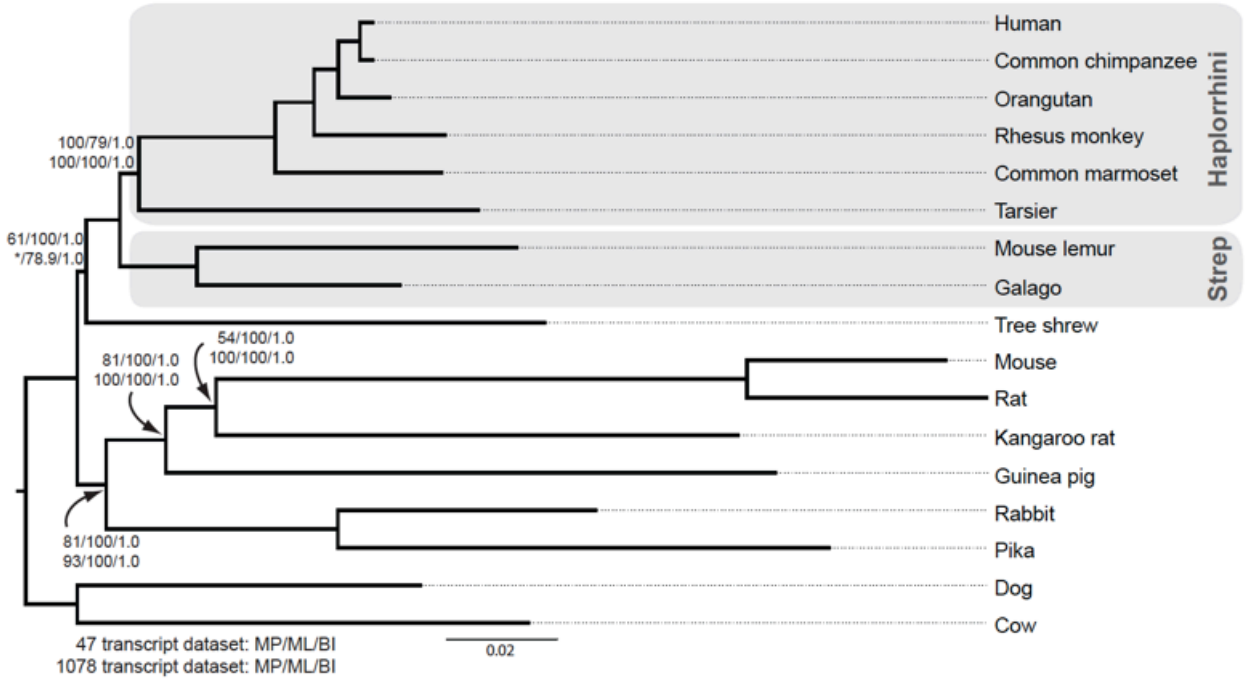


Figure 2.2: Anthropoid phylogenetic tree. Maximum likelihood tree using the 1,078 transcript dataset. % bootstrap and posterior probability scores are shown for both the 47 transcript dataset (top) and the 1,078 transcript dataset (bottom) at all nodes with scores less then 100/100/1.0 (MP/ML/BI). *Maximum parsimony analysis of the 1,078 transcript dataset placed treeshrew (Tupaia belangeri) as sister to the Glires (MP = 99).

Alternative tree typologies exploring the four different hypotheses proposed for the relationships between tarsiers, strepsirrhines, and anthropoids (Figure 2.1) were tested using the Shimodaira-Hasegawa (SH) Topology Test. These analyses showed that the grouping of *Tarsius* with Anthropoidea was preferred over the alternative topologies tested (Figure 2.3). The lnL score for the grouping of tarsiers with anthropoids is significantly better (p<0.001) at explaining the data than the alternate topologies presented (((A,S)T), ((S,T)A), and (A,T,S)).

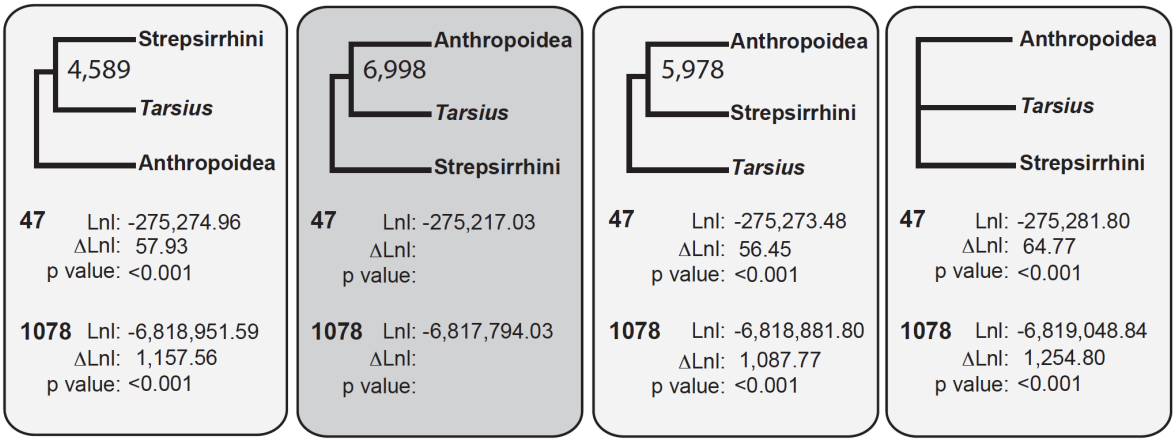


Figure 2.3: Topology test supporting the grouping of Tarsius with Anthropoidea. Shimodaira-Hasegawa topology tests were conducted on both the 47 transcript and the 1,078 transcript datasets and show that the grouping of Tarsius with Anthropoidea (hypothesis B) is a significantly better fit of the data then the other topologies tested ($p<0.001$). The grouping of strepsirrhines with tarsiers is supported by 4,589 substitutions, the grouping of anthropoids and strepsirrhines is supported by 5,978 substitutions, and the grouping of anthropoids with tarsiers is supported by 6,998 substitutions.

Divergence date and rate estimates

The mcmctree divergence date estimations among the 17 taxa used in the phylogenetic analyses are shown in Figure 4. Complete divergence data for every node and every analysis is provided as supplementary material (Table S3). The results from the different calibration models were very similar, with the estimates for each node differing by less than ~1 Ma on average (Table 2.3). Under these different models, the estimate for the most recent common ancestor (MRCA) of Tarsius and Anthropoidea was 68.6 Ma using the bounded fossil calibration method and 67.8 Ma using the probabilistic calibration method, with the credibility intervals for these estimates ranging from 64.9 Ma to 72.7 Ma. The MRCA of Haplorrhini and Strepsirrhini was estimated to be only slightly older and the two different calibration models resulted in estimates of 72.6 Ma and 71.8 Ma respectively. The credibility intervals for these estimates ranged from a lowest estimate of 68.9 Ma to a highest estimate of 76.9 Ma. The estimate for the

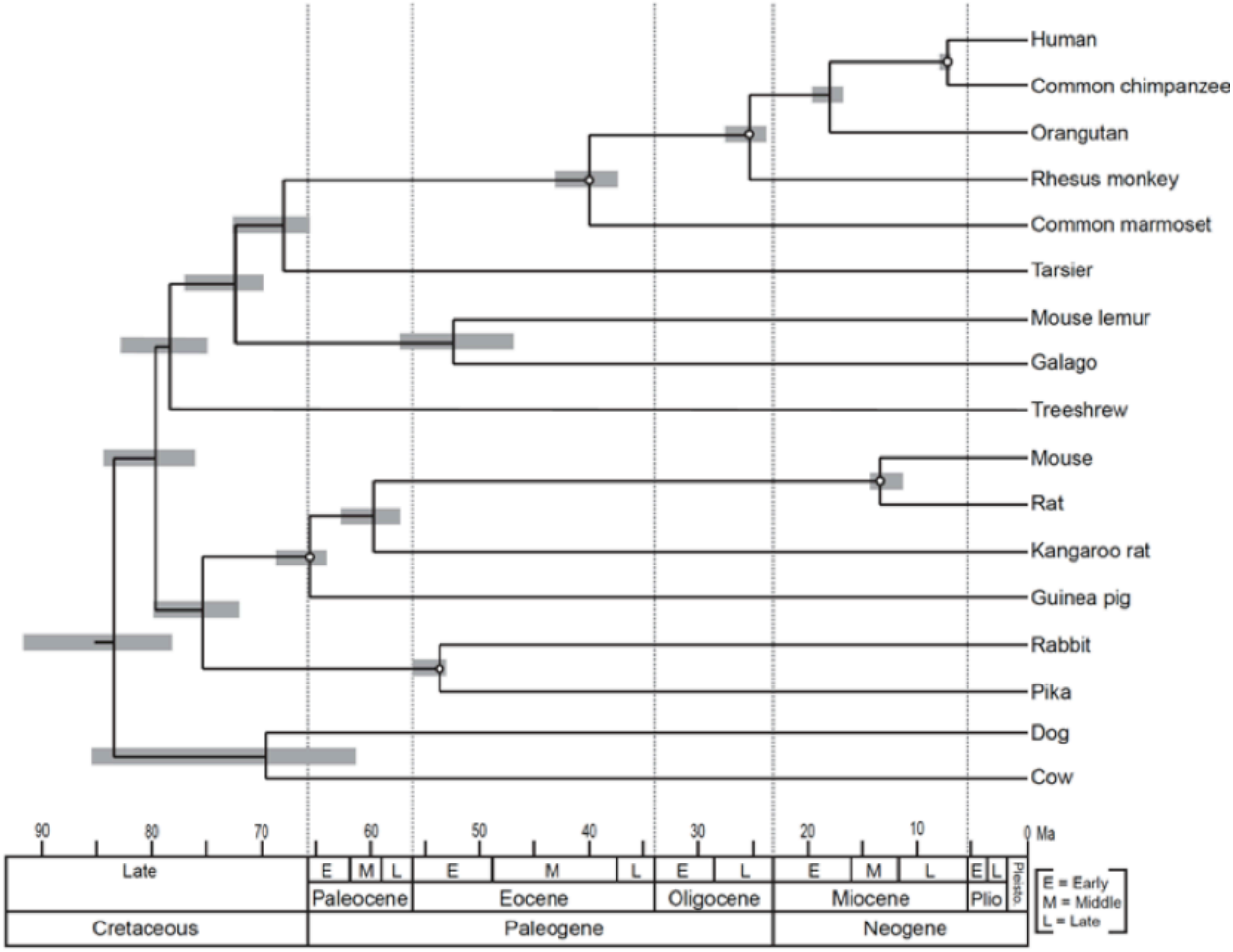


Figure 2.4: Divergence date estimates in Euarchontoglires. The tree shows estimates obtained using the bounded fossil calibrations, shaded bars represent the 95% credibility intervals at each node, and open circles denote nodes in which fossil distributions were used for calibration. Values for the average node dates and credibility intervals are shown in Table 2.3.

Table 2.3

Divergence date estimates and 95% credibility intervals determined for out 17 mammalian taxa using both bounded fossil and probabalistic calibration points in *mcmctree*

	Bounded Fossil Calibrations			Probabilistic Calibrations		
Divergence node	Mean	95% Credibility Interval		Mean	95% Credibility Interval	
Euarchontoglires - Laurasiatheria	83.3	78.1	91.9	82.9	77.8	91.5
Euarchonta - Glires	79.6	76.1	84.4	78.8	75.5	83.3
Scanadentia - Primates	78.2	74.9	82.9	77.4	74.2	81.8
Strepsirrhini - Haplorrhini	72.6	69.6	76.9	71.8	68.9	75.8
Anthropoidea - *Tarsius*	68.6	65.6	72.7	67.8	64.9	71.7
Catarrhini - Platyrrhini	40.0	37.3	43.1	38.0	33.8	42.0
Cercopithecidae - Hominidae	25.4	23.7	27.6	24.3	21.8	26.9
Hominina - Pongina	18.0	16.8	19.6	16.1	13.6	18.7
Homo - Pan	7.2	7.0	7.8	6.2	5.1	7.3
Microcebus - Otolemur	52.4	47.0	57.2	51.2	45.5	56.3
Lagomorpha - Rodentia	75.3	72.1	79.8	74.5	71.5	78.7
Oractolagus - Ochotona	53.7	53.0	56.2	53.6	53.0	56.0
Rodentia	65.6	64.0	68.8	65.3	63.8	68.0
Dipodomys-Muridae	59.7	57.3	62.9	59.5	57.1	62.2
Mus - Rattus	13.4	11.2	14.1	13.4	11.3	14.1
Bos - Canis	69.5	61.2	85.3	69.8	61.2	85.6

MRCA of Microcebus and Otolemur was 52.4 Ma and 51.2 Ma with credibility intervals from 45.5 - 57.2 Ma. The MRCA of Catarrhini and Platyrrhini was estimated at 40.0 Ma and 38.0 Ma with credibility intervals from 33.8 - 43.1 Ma. Within anthropoids the MCRA of Cercopithecidae and Hominidae was estimated at 25.4 Ma and 24.3 Ma with a credibility interval from 21.8 - 27.6 Ma, and the MRCA of Hominina and Pongina (*sensu* (Wildman and Goodman, 2004) was estimated at 18 Ma and 16.1 Ma with a credibility interval from 13.6 - 19.6 Ma. The MRCA of Homo and Pan is estimated at 7.2 Ma and 6.2 Ma using the bounded fossil and probabilistic calibration models respectively, and the credibility interval ranges from 5.1 - 7.8 Ma. Divergence date analysis in BEAST yielded mean divergence dates for the MRCA of tarsiers and anthropoids (58.6 Ma) and the MRCA of haplorrhines and strepsirrhines (64.9 Ma) that are more recent than the mcmctree divergence dates. These dates are consistent with previous estimates (e.g. Goodman et al., 1998). The BEAST results are presented in supplementary Table S4.

Discussion

Here we present a well-supported primate phylogeny using a comparative genomic approach that makes use of the newly assembled *Tarsius syrichta* genome. Maximum parsimony, maximum likelihood, and Bayesian analyses performed on both a 47 transcript and a 1,078 transcript dataset, each comprised of orthologs from 17 mammalian species, demonstrated that *Tarsius* and Anthropoidea are sister taxa, to the exclusion of Strepsirrhini. These phylogenomic data confirm previously published research (Baba et al., 1982; Goodman et al., 1998; Schmitz et al., 2001; Zietkiewicz et al., 1999). The topology tests, as well as branch support values, decisively support a monophyletic haplorrhine clade and reject the three alternative hypotheses for the

relationship among the three major extant primate clades. Indeed, 6,998 nucleotide site synapomorphies support the haplorrhine clade. The tree that groups strepsirrhines with tarsiers is supported by only 4,589 substitutions while an anthropoid and strepsirrhine grouping is supported by 5,978 substitutions.

A genomic approach

In the present study, we were able to sample a wide range of genes, rather than relying on the phylogenetic signal derived from a single locus (e.g. the cytochrome b gene). This strategy has enabled us to obtain a greater degree of statistical power than has been available in previous studies that have examined the relationships among primate taxa. Utilization of whole genome assemblies allowed us to generate a 1.26 Mbp multiple sequence alignment across 17 species that covered a large number of loci. Even though several of the genome assemblies used in the study, including that of *Tarsius syrichta*, are represented by only ≤ 2 fold genome coverage (Table 2) and thus may contain sequencing errors and assembly gaps (Green, 2007), the larger number of loci examined and the removal of clearly misaligned and inaccurate sequence from the multispecies alignment gives us confidence in the accuracy of our phylogenetic conclusions. Moreover, our approach is quite conservative because we included only 1:1 orthologs in the study, i.e., we removed genes that had duplicated during eutherian phylogeny. Although removing such genes reduced our pool of data, it provided us with high quality datasets significantly larger than those previously used to address the question of tarsier phylogeny.

Divergence dating

Both methods of divergence dating produced roughly comparable date estimates; however, the majority of divergence dates estimated by BEAST were more recent than those obtained with mcmctree. These more recent estimates are more similar to dates previously published (Goodman et al., 1998). Using mcmctree the origin of crown haplorrhines is estimated at approximately 68 Ma and, when considering credibility intervals, ranges from 64.9 Ma to 72.7 Ma. This places the origin of haplorrhines either in the Cretaceous or very close to the KT boundary. The origin of the haplorrhine lineage occurred very shortly after the origin of primates, which our clock analyses placed at approximately 72 Ma. This date suggests that primates not only had their origin in the Cretaceous, but that the order may have further diversified in this period. Regardless of the exact dates, the origin of crown primates and haplorrhines occurred in rapid succession. The molecular clock analyses suggest that the age of the haplorrhine node is approximately 94% of the age of the crown primate node. Given the relatively small time interval between the two nodes, the difficulty in determining the phylogenetic position of tarsiers is not surprising. The haplorrhine-strepsirrhine divergence date presented here of 68 (64.9 - 72.2) Ma, falls between previously published divergence dates that range from 58 Ma (Gingerich and Uhen, 1994; Goodman et al., 1999; Goodman et al., 1998) to 74 Ma (Seo et al., 2004). Our findings suggest a Cretaceous origin for the haplorrhine clade and, as noted previously, a Cretaceous origin of Haplorrhini that precedes the first appearance of Primates and Euprimates in the fossil record (Miller et al., 2005). However, all of our primate calibrations were from fossils dated to the late Eocene or earlier; thus, if early primates

had a different substitution rate than occurred in the past 35 million years, we would potentially infer an incorrect time of origin for the crown haplorrhine and primate nodes. We attempted to solve this problem by including calibrations outside of the primate clade (i.e., within rodents). Specifically, the conservative upper estimate for the crown rodent origin (<65 Ma), led to relatively rapid estimates for the speed of the molecular clock within the early Euarchontoglires lineages. In analyses that did not set an upper bound for the crown rodent divergence, even earlier estimates for primate and haplorrhine origins were estimated (results not shown). The discord between the fossil and molecular timescales of primates is well known (Steiper and Young, 2008), and this issue will remain apparent until the development of novel molecular clock methodologies or discovery of fossil evidence to support or reject our estimated dates.

Tarsier diversity

First described in 1706 by J. G. Camel, the genus *Tarsius* now contains at least eight species (*T. bancanus*, *T. dentatus*, *T. lariang*, *T. pelengensis*, *T. tarsier*, *T. sangirensis*, *T. pumilus*, and *T. syrichta*) (Groves, 2005; Merker and Groves, 2006). A population of tarsiers from the Indonesian Siau Island has been recognized as being distinct from the eight recognized species, but has yet to be named as a new species (Mittermeier et al., 2007). Several fossil tarsiers have been identified throughout Asia. However, *Tarsius eocaenus* is considered an extinct tarsier species belonging to the extant genus. This fossil was recovered in the fissure-filings of Jiangsu Province, China, a richly fossiliferous geological region dating back to the middle Eocene (~45 Ma) (Beard et al., 1994). Including *Tarsius eocaenus* in the genus *Tarsius* has deepened the genus such that it encompasses 45 million years of evolution. Our divergence estimates

suggest that *T. eocaenus* lived 20 million years after the origin of the tarsiiform lineage. Such deep divergence of a genus is rarely reported, although *Solenodon* and elephant shrew both show intra-genus divergence dates >20 Ma (Roca et al., 2004; Tabuce et al., 2007).

Primates are members of Euarchontoglires, a superordinal clade that includes all extant rodents, lagomorphs, scandentians, dermopterans, and primates. Of these orders, all but primates are of unambiguous Eurasian origin, and several hypotheses exist on the biogeographic origin of primates (Miller et al., 2005). Paleogene anthropoids are known both from Asia and Africa (Williams et al., 2010). Similarly, tarsiers have been described from the Paleogene of Asia and from the early Neogene of Africa. Biogeographically, this is remarkable, as it suggests either 1) the dispersal of anthropoids and tarsiers across the Tethys Sea, or 2) a very early divergence of haplorrhines followed by vicariance via continental drift. Our divergence date estimates provide more support for the first scenario because vicariance due to the breakup of Gondwana from Laurasia, and the subsequent split between Africa and the Indo-Madagascan land mass, occurred at least 25 million years before the estimated time of the last common ancestor of haplorrhines.

Tarsiers, anthropoid's closest relatives

The current study demonstrates that tarsiers are the closest living relatives of the extant anthropoid primates. Although tarsiers and strepsirrhines are often gradistically considered more "primitive" primates, and thus grouped together, we reject the idea that these primates are "primitive" and our data very strongly support a monophyletic haplorrhine clade. Indeed, the branch lengths depicted in Figure 2.2 indicate that there

has been more genomic evolution on the lineage leading to *Tarsius* than has been seen in any sampled anthropoid lineage. We propose that our data strongly reject the division of “lower” and “higher” primates, and instead, we argue that three main extant lineages have persisted since the time of the last common ancestor of primates. The molecular data suggest anthropoids have fewer nucleotide substitutions than either strepsirrhines or tarsiiforms (Figure 2.2), and as such, anthropoids appear to have undergone the least amount of molecular evolution of these three primate groups. Within Anthropoidea, the hominids show the slowest rate of nucleotide substitution confirming previous findings (Bailey et al., 1991; Bailey et al., 1992; Elango et al., 2006). However, we note that the present study is primarily limited to coding sequences, and we have not explored rates of evolution in other regions of the genome. Thus, future work is required to determine just how much genomic evolution has occurred in the different primate lineages.

Distinct traits of the tarsier genus

A more exact primate phylogeny enables us to more accurately reconstruct the ancestral genotypes and phenotypes of stem Haplorrhini and Anthropoidea. There has been considerable discussion and debate concerning the nature of the morphological and behavioral characteristics of tarsiers and their evolutionary origins. As an example, nocturnal behavior is often considered a synapomorphy, or a shared derived trait, between tarsiers and strepsirrhines (Napier and Napier, 1985). Although tarsiers and some strepsirrhines have morphological adaptations that support a nocturnal lifestyle, the specific adaptations, such as lack of a tapetum lucidum, increase in number of photoreceptors, and increase in overall eye size are themselves not shared between the

two groups. Indeed, it has been proposed that the nocturnal activity pattern exhibited by tarsiers emerged after the origin of the haplorrhine clade (Martin, 1990). Most nocturnal mammals (e.g., opossum, cat, hedgehog, and rabbit) have a tapetum lucidum that reflects visible light back through the retina increasing the amount of light available to photoreceptors (Ollivier et al., 2004). This adaptation improves vision in low-light conditions and allows for night vision. Tarsiers, however, lack a tapetum lucidum and instead have significantly larger eyes (with many more photoreceptors) that improve vision in low-light conditions (Hendrickson et al., 2000; Martin, 1973; Pocock, 1918). Tarsiers also have the largest eyes relative to body size of all extant primates (Schwartz, 2003). Similar adaptations for night vision, such as an increase in the number of photoreceptors, are also seen in the owl monkey (*Aotus*). These two species are the only nocturnal haplorrhines and appear to have homoplastically evolved from a diurnal to a nocturnal lifestyle. Based on the phylogenetic data presented here, as well as the distinct optical morphology of the tarsiers, we agree with previous research suggesting that tarsiers have made a secondary return from a diurnal to a nocturnal lifestyle (Martin, 1990) and that nocturnal behavior is a secondarily derived homoplastic feature in *Tarsius* and *Aotus*.

Another trait mistakenly categorized as derived among tarsiers and anthropoids is the level of intimacy of the placenta at the maternal fetal interface (e.g., Luckett, 1974). The highly invasive hemochorial placenta in both anthropoid primates and tarsiers was thought to have originated on the haplorrhine stem and had been used to support the hypothesis of a monophyletic Haplorrhini. It is now known that the placenta of the ancestral eutherian mammal was hemochorial (Elliot and Crespi, 2009; Wildman

et al., 2006). The assumption that the hemochorial placenta represents a derived haplorrhine trait neglects to consider mammalian phylogeny in broader context. With the description of the Euarchontoglires clade, it became apparent that the strepsirrhini possess the derived epitheliochorial type of placenta interface. Rodents, lagomomorphs, dermopterans, and haplorrhine primates all possess the ancestral hemochorial placenta interface, and the hemochorial interface is found outside of Euarchontoglires in the other three placental mammal clades (i.e., Afrotheria, Xenarthra, and Laurasiatheria). In order to reconstruct the evolutionary history of the remaining features listed in Table 2.1, it will be necessary to examine the character state distributions for these characters in groups outside of Primates, rather than to assume that strepsirrhines and/or prosimians posess the ancestral state for these characters.

Conclusion

The relationship of tarsiers to anthropoids and strepsirrhines has been a controversial topic for more than a century (Martin, 1990; Wright et al., 2003). The current study demonstrates strongly that tarsiers belong to the haplorrhine clade. These creatures have undergone a greater amount of protein-coding nucleotide substitution during their descent from the most recent common ancestor of extant haplorrhines than have their sister taxon, the Anthropoidea. We have estimated 1) that primates emerged as a total group by 78.2 Ma, 2) that the last common ancestor of extant primates lived approximately 72.6 Ma, and 3) that haplorrhines emerged by 68.6 Ma. Although our phylogeny is strongly supported by the data, the inclusion of additional tarsier species in future studies would add to our findings and provide insight into evolutionary events occurring within the genus. If genomic data from other tarsier species become available

it would be very interesting to revisit the data in order to infer the time of origin of extant tarsiers.

Acknowledgements

The research was funded by NSF grants BCS-0751508 and BCS-0827549. Natalie Jameson is funded by NSF Doctoral Dissertation Improvement Grant BCS-1061370. We acknowledge The Genome Center at Washington University as the source of the unpublished tarsier genomic data. We thank the Wayne State University (Detroit, MI, USA) high performance grid for access to computing resources. Our dear colleague and coauthor, Dr. Morris Goodman passed away during the revisions of this manuscript. We are ever grateful for his intellectual honesty and dedication to the study of primate phylogenomics.

CHAPTER 3

Development and annotation of shotgun sequence libraries from New World monkeys

This chapter has been published as:
Jameson, N.M., Xu, K., Yi, S.V., Wildman, D.E., 2012. Development and annotation of shotgun sequence libraries from New World monkeys. Molecular Ecology Resources 12, 950-955.

Summary

The draft genome sequences of several primates are available, providing insights into evolutionary and anthropological research. However, genomic resources from New World monkeys are conspicuously lacking. To date, the genomes of only two platyrrhine species, the common marmoset and the Bolivian squirrel monkey have been fully sequenced. This is especially limiting for comparative genomics research, considering that New World monkeys are the most speciose primate group, and platyrrhine genetic diversity is comparable to that of the catarrhines (i.e. apes and Old World monkeys).

Here we present the generation and annotation of numerous sequence reads from the genomes of Spider monkey (*Ateles belzebuth*), Owl monkey (*Aotus lemurinus*), and Uakari (*Cacajao calvus*) representing the three platyrrhine families, Atelidae, Cebidae, and Pitheciidae, respectively. These sequencing reads were developed from gDNA shotgun libraries containing over 3,000 individual sequences with an average length of 726 bps. Of these sequences 1,220 contain less than 20% repeats, and thus are potentially highly useful phylogenetic markers for other platyrrhine species. Among them, a large number of sequencing reads were found to match unique regions within

the human (2,462 sequences) and the marmoset (2,829 sequences) genomes. In particular, the majority of these sequencing reads are from putatively neutrally evolving intergenic regions. Thus, they are likely to be highly informative for inferring neutral evolutionary patterns and genomic evolution for other New World monkeys.

Introduction

New World monkeys (Primates: Platyrrhini) are the clade of primates that inhabit South and Central America (Groves, 2005). The ancestral platyrrhines dispersed to South America between 40 and 26 million years ago thus splitting crown anthropoid primates into two distinct groups, 1) Platyrrhini, and 2) the Catarrhini, comprised of Old World monkeys, and apes including humans (Ciochon and Chiarelli, 1980a; Houle, 1999). Upon dispersal, New World monkeys began a period of rapid diversification and divergence filling new environmental niches. This period of rapid divergence has resulted in the most speciose extant primate group, including three families, 15 genera, and 126 species (Goodman et al., 2005; Groves, 2001, 2005; Opazo et al., 2006; Wildman and Goodman, 2004). The pronounced variation among species can be seen in the form of both morphological and molecular differences. For example there is an approximately 100-fold weight difference between the 110g pygmy marmoset (*Callithrix pygmaea*) and the 11.4kg Mexican howler monkey (*Alouatta palliata*). Moreover, platyrrhines include the only nocturnal anthropoid primate genus, *Aotus*; and platyrrhines range in chromosome number from 2n=16 in the black titi monkey (*Callicebus lugens*) to 2n=58 in Hershkovitz's Owl monkey (*Aotus hershkovitzi*) (Stanyon et al., 2003; Torres et al., 1998).

The morphology, behavior and ecology of New World monkeys has been widely studied and is well documented (i.e. Ford, 1986; Hershkovitz, 1977). This knowledge base has informed studies in the fields of phylogenetics, immunology, anthropology, taxonomy and population dynamics (Babb et al., 2010; Ford, 1986; Genain and Hauser, 2001; Hartwig et al., 2011; Holmes et al., 1966; Kay, 1990; Marroig and Cheverud, 2001; Opazo et al., 2006; Organ et al., 2009; Papper et al., 2009; Rosenberger, 1984; Terhune, 2011; Wildman et al., 2009). However, the availability of genetic data for New World monkeys remains limited. Although New World monkeys account for roughly one third of all primate species, until very recently only one of the 10 publicly available fully sequenced primate genomes was from a platyrrhine. In addition to the annotated genome of the common marmoset (*Callithrix jacchus*) the draft genome assembly of the Bolivian squirrel monkey (*Saimiri boliviensis*) has recently been released. Due to the limited amount of molecular data available, many New World monkey studies have instead relied on molecular markers designed based on human sequence data (Clisson et al., 2000; Witte and Rogers, 1999). The majority of primers designed using human genome sequences fail when employed on New World monkeys because of the long evolutionary distance between them and humans. Thus, it is imperative to develop molecular markers from New World monkey genomes.

In order to establish a database of high quality genetic data we have developed genomic DNA shotgun libraries from representatives of each of the three extant New World monkey families; Atelidae, Cebidae and Pitheciidae (Figure 3.1). The white-fronted spider monkey (*Ateles belzebuth*), a member of the Atelidae inhabits the north-western Amazon and possesses the unique feature of a prehensile tail, a trait shared

only by the Atelids and the capuchin monkey. The grey-bellied night monkey (*Aotus lemurinus*) inhabits the northern tropical Andes, one of only 11 nocturnal anthropoid primates species, and is a member of the Cebidae (Morales-Jimenez and de la Torre, 2008). The bald uakari (*Cacajao calvus*) inhabits the western Amazon, and represents one of the four extant genera that comprise the Pitheciidae (Veiga et al., 2008). The inclusion of *Ateles belzebuth, Aotus lemurinus, and Cacajao calvus* in our shotgun libraries allowed us to capture a wide breadth of genetic diversity that exists through the New World monkey families.

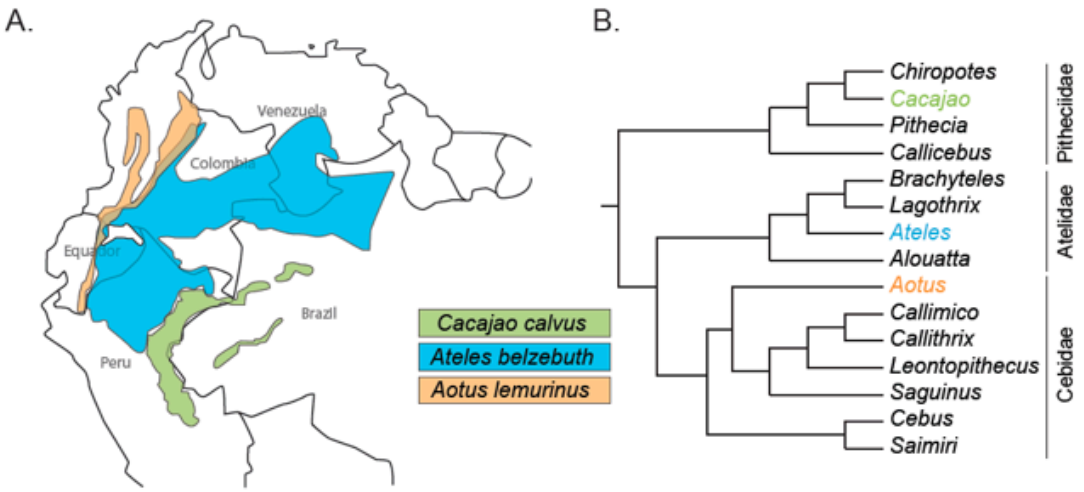


Figure 3.1: A. Partial map of South America showing the distribution of the ranges of the three species of New World monkey sequenced in this study (distribution based on those of the IUCN Red List, Boubi et al. 2008; Morales-Jimenez & de la Torre 2008; Veiga et al. 2008). B. Phylogenetic tree adapted from Wildman et al. 2009 showing the relationships of all New World monkey genera and the separation of the three New World monkey families.

Here we present a molecular resource that includes 3,154 individual genomic sequencing reads containing over 2.3 million nucleotides and is, to date, the largest single molecular resource for these species. A detailed annotation of all sequencing reads and assembled reads as well as alignments of both individual and assembled reads to the human and marmoset genomes are presented in the supplemental data files of this manuscript. Individual sequencing reads are all available through NCBI Genbank (JQ930030 - JQ933105 and JQ963345-JQ963421) and assembled sequence contigs are available in the supplemental data files. Supplemental data files associated with this manuscript are available through the following web link: http://onlinelibrary.wiley.com/doi/10.1111/j.1755-0998.2012. 03162.x/suppinfo.

Materials and Methods

Tissue collection and DNA isolation

Whole blood samples from *Cacajao calvus* and *Ateles belzebuth* were obtained from the Centro de Primatologia de Rio de Janeiro and the Brazil Centro Nacional de Primates, respectively. DNA was isolated from the blood samples using the Eppendorf (Hamburg, Germany) gDNA blood mini kit. DNA was also isolated from an *Aotus lemurinus* liver sample using the Qiagen Inc. (Valencia, CA) DNeasy kit.

Generation of genomic shotgun libraries

DNA was sheared for 90 seconds at 10 psi using a nebulizer. Randomly cut DNA was first blunt-end repaired and dephosphorylated, then ligated into pCR® 4Blunt-TOPO® vectors (part of the TOPO® Shotgun Subcloning Kit, Invitrogen). Ligation products were transformed in chemically competent *E. coli* cells. Transformants were placed on selective plates and incubated overnight at 37°C. Individual colonies were

picked and grown overnight in LB media with ampicillin resistance. Vectors with insert were isolated using the Qiaprep Miniprep procedure (Qiagen, Valencia, CA). These inserts were then sequenced in both directions with Big Dye v.3.1 (Applied Biosystems, Foster City, CA), using T3 and T7 universal primers on the ABI 3730xl genetic analyzer. This resulted in a total of 3,286 individual sequences. The process of generating and annotating the shotgun libraries is detailed in the workflow in Figure 3.2.

Sequence read analysis

Vector and primer sequences were trimmed from the resulting sequencing reads and failed cycle sequencing reactions (sequences containing >300 N's) were discarded from further analysis. This resulted in 3,154 sequences remaining for analysis. These sequences have been deposited into Genbank (Accession IDs: JQ930030 - JQ933105 and JQ963345-JQ963421). Repetitive elements within sequences were detected, annotated, and masked using RepeatMasker open-3.0 (Smit, 1996-2004). The shotgun sequences were separately queried in BLAST (Altschul et al., 1990) using the blastn algorithm against the human (GRCh37/hg19) and marmoset (WUGSC3.2 /calJac3) genomes. For both searches the criteria for selecting hits were a) E-value $<1.0E^{-50}$ and b) HSPs must match more than 50% of the query sequence. The human genome annotation was used to determine the location of the BLAST hits. To determine orthology between the human and marmoset genomes, we used the reciprocal BLAST approach. Pairwise alignments against both human and marmoset were generated using T-coffee (Version_8.93; Notredame et al., 2000) for those shotgun library sequences having significant blast hits and containing less than 30% repeats. The alignments of the sequencing reads to the human and marmoset genomes are available

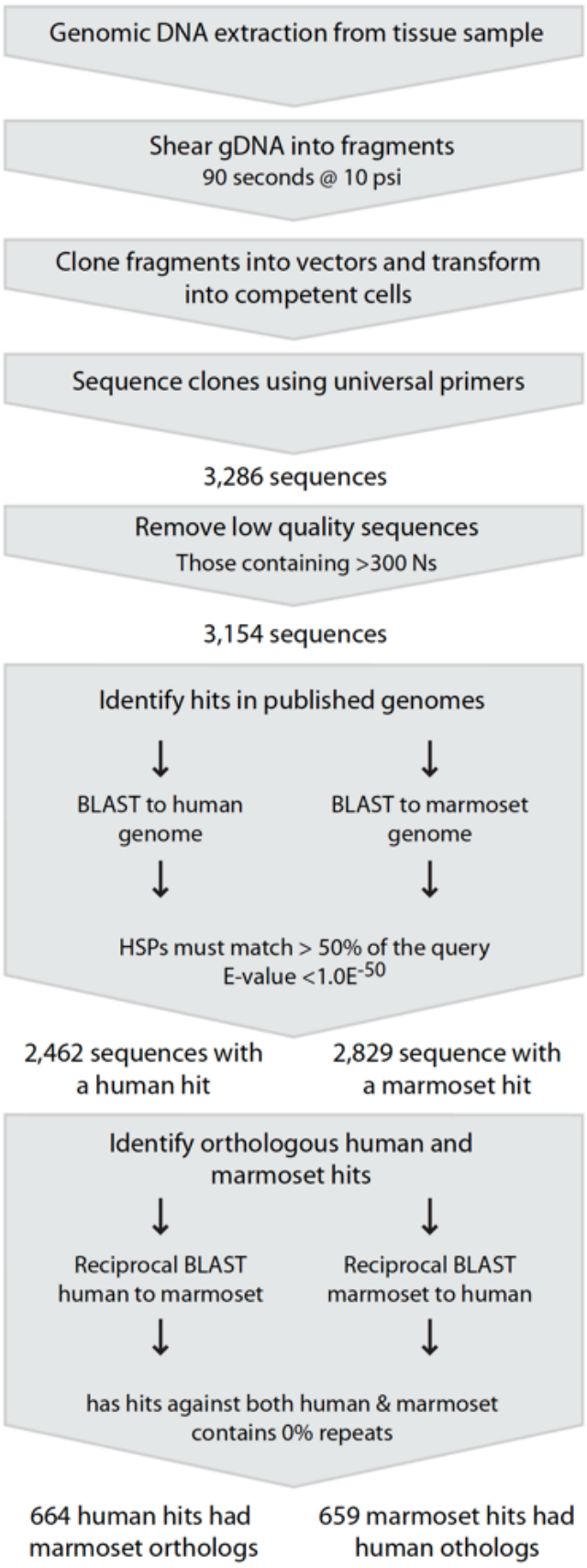


Figure 3.2: Workflow showing the steps involved in the shotgun library construction, sequencing, and annotation.

in the online Supplemental Data files (Sequencing_Reads_Human_Alignmants.txt and Sequencing_Reads_Marmoset _Alignments.txt). Sequence reads containing less than 20% repeats were assembled into contigs using Phrap (Delson and Rosenberger, 1980; Gingerich, 1980) while maintaining the default parameters. These contigs were then queried in BLAST against the human and marmoset genomes as described above. The alignments of the contigs to the human and marmoset genomes and the sequences of all assembled contigs are available in online Supplemental Data files (Assembled _Contig_Human_Alignments.txt, Assembled_Contig_Marmoset_Alignments.txt, and _Contig_Sequences.txt).

Results

Sequence Composition

A total of 3,286 sequencing reads were obtained from the shotgun libraries. After removing low quality sequences the dataset contained 3,154 sequence reads, 1,272 from the *Ateles* library, 1,045 from the *Cacajao* library, and 837 from the *Aotus* library. This dataset consists of 2,282,823 bps, the shortest and longest sequences measure 52 bp and 961 bp respectively, and the mean length of sequences in the dataset is 726 bp (Table 3.1). The GC content of all sequences from the *Ateles* library was 41.8%, 39.5% from the *Cacajao* library sequences and 40.6% from the *Aotus* library sequences. Of the 3,154 sequences, 813 were found to be free of repetitive elements, 1,220 sequence reads contained <20% repeats and 42 sequence reads contained 100% repeats. Of the repetitive sequences, 36.12% were LINEs, 33.66% were SINEs and 15.32% were LTR elements (Table S5) as identified by RepeatMasker open-3.0 (Smit 1996-2004). The

abundance and composition of different transposable elements are similar to what's been observed in other primate genomes (Consortium 2005; Lander et al. 2001).

Table 3.1
Shotgun library composition

		Individual sequence lengths					
Shotgun Libraries	Sequences	Minimum	Maximum	Mean	Median	Basepairs	GC %
Aotus lemurinus	837	133	957	737	793	616,971	40.6
Ateles belzebuth	1272	52	836	697	717	887,261	41.8
Cacajao calvus	1045	52	961	745	754	778,591	39.5
Combined	3154	52	961	726	754.6	2,282,823	

Sequence Annotation

BLAST searches identified 2,462 sequencing reads with at least one hit in the human genome, 1,085 of the sequencing reads contained <20% repeats and 704 were free of repetitive elements. In addition 2,829 of the sequence reads produced at least one hit against the marmoset genome, 1,153 of which contained <20% repeats and 735 were free of repetitive elements. As anticipated, the BLAST query against the marmoset genome identified more regions of sequence identity than the query against the human genome. Sequencing reads that produced hits against the human genome covered all chromosomes. Human chromosome 3 contained the most sequencing reads and chromosome 21 contained the least (293 and 28, respectively); see Figure S2 for more details. With the exception of a few outliers, the number of sequencing reads per chromosome is correlated with the length of the chromosomes (R^2=0.74, p=0.016). Using the human genome annotation, 57% of the BLAST hits are in intergenic regions, 38% are in introns, 2.2% are within 2 kb upstream or downstream of transcriptional start sites, 0.2% are inside exons, and the remaining span across two or more of the regions specified above. These frequencies are in accord with the estimated portions of different

genomic regions in primate genomes (Consortium 2005; Lander et al. 2001). Thus, the sequencing reads developed in this study appear to be well representative of different genomic regions. Full annotation details for each sequencing read are available in the online Supplemental data files (Sequencing_Read _Annotation.xls).

Orthologous sequences

Among all sequencing reads, 692 included no repeats and had hits in both the human and the marmoset genomes. We also tested the presence of putative single copy orthologs between human and marmoset genomes among these sequencing reads. We found that 92.3% of the sequencing reads (639 pairs) were single copy orthologs between human and marmoset genomes, by reciprocal BLAST hit searches between the human and marmoset genomes. Thus our method is highly efficient in identifying orthologous sequencing reads. Because genomic coverage is relatively low (i.e 0.01% of the genome was sequenced in each species), there were no sequencing reads that were orthologous between the three shotgun libraries of *Aotus*, *Cacajao* and *Ateles*.

Sequence Assembly

The 1,220 sequencing reads containing less than 20% repeats were assembled into contigs using Phrap (Delson and Rosenberger, 1980; Gingerich, 1980). The assembly produced a total of 186 contigs that contain 66% of the sequencing reads; this brought the total number of unique genomic regions identified in this study to 658 (158 from *Aotus lemurinus*, 250 from *Ateles belzebuth*, and 250 from *Cacajao calvus*). Full annotation details for each assembled contig are shown in the online Supplemental data file (Assembled_Contig _Annotation.xls).

Discussion

The dataset presented here is the largest single molecular resource for these species, and should prove to be useful for a variety of downstream applications. None of the genomes included in the shotgun libraries (*Aotus lemurinus, Ateles belzebuth or Cacajao calvus)* are scheduled for whole genome sequencing in the near future, and currently New World monkey draft genome data are limited to the common marmoset (*Callithrix jacchus*) and the Bolivian squirrel monkey (*Saimiri boliviensis*). Although some molecular genetic work has been done on the species included in the present study (i.e. Menezes et al., 2010; Opazo et al., 2006; Perelman et al., 2011; Wildman et al., 2009), the resources are still rather limited. For example, there are only 402 Entrez nucleotide entries belonging to these species in NCBI (www.ncbi.nlm.nih.gov/, as of April 2012). The addition of this large-scale molecular resource works to fill a gap in the molecular knowledge base available within New World monkeys.

Our annotations determined these sequences to exist in a variety of genic and intergenic regions. However, given the limited depth of our sequence coverage we can not rule out the possibility that those sequences determined to be located in genes are not in pseudogenes or another member of the gene family. Due to the diversity of genomic regions captured by shotgun libraries they present a good source of variety of molecular markers. As an example previous and ongoing research by our group has resulted in the development of non-genic markers, which were used for phylogenetic analysis, developed from similar shotgun libraries (Wildman et al., 2009). In particular, markers generated from putatively neutral regions were highly useful to resolve phylogenetic relations (Wildman et al., 2009) as well as to determine molecular rate

variation (Peng et al., 2009). In addition to the non-genic sequences, these data provide a good resource for the development of various types of molecular markers that could be used in phylogenetic, population, or conservation studies. As the majority of these sequences were also found in either the human or marmoset genome assemblies, these markers have and will continue to be a great resource for comparative genomic study approaches.

We have made this dataset available in anticipation of its use for various downstream applications such as microsatellite marker development and phylogenetic markers (coding, non-coding or non-genic). It provides a valuable contribution to the genetic tools existing for the study of all New World monkey species as well as for other primates.

Acknowledgments

This research was funded by National Science Foundation Grants (BCS-0751508) to D.W. and (BCS-0751481) to S.Y. Natalie Jameson is funded by NSF Doctoral Dissertation Improvement Grant BCS-1061370. We thank The Centro de Primatologia de Rio de Janeiro, Centro Nacional de Primatas, Belém, Pará, Brazil, and Dr. Caro-Beth Stewart for providing samples.

CHAPTER 4

The tempo and mode of New World monkey evolution and biogeography in the context of phylogenomic analysis

This chapter is under review at Molecular Phylogenetics and Evolution:

Jameson, N.M., Yi, S.V., Xu, K., Wildman, D.E.,. The tempo and mode of New World monkey evolution and biogeography in the context of phylogenomic analysis. Molecular Phylogenetics and Evolution, In Review.

Summary

The development and evolution of an organism is heavily influenced by its environment. Thus, understanding the historical biogeography of a clade can provide insights into their evolutionary history, adaptations and trade-offs realized throughout time. In the present study we have taken a phylogenomic approach to infer New World monkey phylogeny, upon which we have reconstructed the biogeographic history of extant platyrrhines. In order to generate sufficient phylogenetic signal within the New World monkey clade, we carried out a large-scale phylogenetic analysis of approximately 40 kb of non-genic genomic DNA sequence in each of 36 New World monkey species. Maximum parsimony, maximum likelihood and Bayesian inference analysis all converged on a single optimal tree topology. Using the phylogenetic framework, divergence date estimates and extant species ranges, we reconstructed the ancestral geographic ranges using the RASP package (S-DIVA and BBM). The

ancestral area reconstruction describes the geographic locations of the last common ancestor of extant platyrrhines and provides insight into key biogeographic events occurring during platyrrhine diversification.

Introduction

The oldest primate fossil found in South America is dated to 26 million years ago (Fleagle and Tejedor, 2002; MacFadden, 1990), suggesting the dispersal of primates to the Neotropics occurred sometime before then. This dispersal likely occurred during the time period when South America was an island continent. The geologic isolation of South America began in the late Cretaceous after the breakup of Gondwana and ended when the continent joined with North America via the Isthmus of Panama in the Pliocene (Ciochon and Chiarelli, 1980a). Upon dispersal to the Neotropics, primates underwent a relatively rapid divergence, filling various novel environmental niches, potentially facilitated by the lack of competition for resources from native species. As a result of this dispersal, neotropical primates are monophyletic, and make up all species included within the clade Platyrrhini. While taxonomic schemes vary, some classifications suggest that today the clade consists of 128 species, 15 genera and 3 families (Groves, 2005; Wildman and Goodman, 2004) that have expansive ranges, and live in a variety of habitats throughout the Neotropics and subtropics from 29° South latitude to 22° North latitude (IUCN, 2012; Kay et al., 2012).

It is unclear precisely how and when primates arrived in South America. Studies have hypothesized both African and Asian origins for platyrrhines and suggested terrestrial routes through North America and Antarctica and aquatic or island-hopping routes directly from Africa (Ciochon and Chiarelli, 1980a). Our current knowledge of the

primate fossil record provides few clues to the route taken; however, research in paleoclimate and paleogeography has provided some insight into the feasibility of various scenarios (e.g. Houle, 1999). Possibly due to the harsh environment of Amazonia, only a few fossil deposits have been identified in South America (Rosenberger, 2002). Many extinct primate specimens have been identified from these rich fossil reserves, however due to the paucity of geographic locales, it is difficult to identify patterns of migration throughout the continent. Taking a historical biogeography approach, we reconstructed the past movements of New World monkeys (NWMs) beginning with their current geographic ranges (Ciochon and Chiarelli, 1980a). Biogeographic analyses integrate distributional, phylogenetic, molecular and paleontological data into a comprehensive study of the patterns of species evolution and the historical changes that shaped them (Crisci, 2001). An understanding of the biogeographic history of NWMs can be achieved by tracing their current distribution, as well as the history of their dispersal and vicariant events, on a well established phylogenetic framework. This will then provide insight into external forces such as the geophysical environment and climatic changes that in conjunction with speciation affected the evolutionary histories of these animals.

Although heavily studied, aspects of the phylogeny of extant platyrrhines has remained ambiguous. Relationships within NWMs that have historically been difficult to resolve include the relationships among the three generally recognized extant NWM families (Cebidae, Atelidae, Pitheciidae) and the arrangement of three monophyletic clades within Cebidae: Callitrichinae, Cebinae and *Aotus* (Canavez et al., 1999; Harada et al., 1995; Porter et al., 1997b; Schneider et al., 2001). The lack of sufficient data is

often a key contributor to phylogenetic uncertainty. This can be somewhat remedied by increasing the number of molecular and/or morphological characters included in the analysis. However certain characteristics intrinsic to a clade's evolutionary history, such as rapid divergence among taxa, limit the amount of phylogenetically informative characters available for study.

Closely spaced speciation events result in short phylogenetic branches, which contain few diagnostic character states. This decreased amount of phylogenetic signal is evidenced by lower branch support scores and greater incongruence among gene trees (Philippe et al., 2011; Wiens et al., 2008). Long branches, however, carry their own set of limitations, such as long-branch attraction (Felsenstein, 1978, 2004; Huelsenbeck, 1995). This condition is exacerbated when rapid speciation events occur alongside long branches, as the limited phylogenetic signal on the short branches is unable to overcome the natural tendency toward long-branch attraction. Lastly gene tree incongruence, due to incomplete lineage sorting, is also intensified by the presence of closely spaced speciation events (Maddison, 1997). Much of the difficulty encountered in resolving NWM phylogeny is due to the lasting effects of their rapid radiation, i.e. short branch lengths. In order to overcome these limitations an alternative type of data may be more successful in resolving such a phylogeny. Studies have shown that the rate of evolution in non-genic DNA is in fact higher than that seen in genic regions, including introns and pseudogenes, indicating a lack of functional constraints and selection pressures on the non-genic regions (Chen and Li, 2001; Ellegren et al., 2003). Phylogenetic studies utilizing non-genic markers have been successful in reconstructing the relationships among a sample of primates illustrating

the phylogenetic usefulness of these novel molecular markers (Peng et al., 2009; Wildman et al., 2009).

Here we infer the phylogenetic relationships within NWMs using non-genic molecular markers. With this framework in place we present an ancestral area reconstruction based on the estimated divergence times and the ranges of extant platyrrhine species. This reconstruction describes the timing and location of migration events, providing insight into the tempo and mode of primate evolution in the Neotropics.

Materials and Methods

Taxon sampling and DNA isolation

We analyzed the phylogenomic history of 40 species, 36 platyrrhines and four non-platyrrhine primate outgroups (Table 4.1). Genomic DNA was isolated from either whole blood (Eppendorf gDNA Blood Mini kit, Hamburg Germany) or tissue (Qiagen DNeasy kit, Valencia CA). Because the DNA samples were in most cases from threatened or endangered species, we applied whole genome amplification (WGA) (Qiagen REPLI-g Midi kit) to conserve valuable DNA stocks. The genomes were amplified by multiple displacement amplification (MDA), a process that uses rolling circle amplification to amplify genomic DNA in segments >10 kb long. In comparison with PCR based WGA methods, MDA has been proven to introduce less amplification biases and have more concordance (99.7%) with unamplified DNA templates (Hughes et al., 2005; Lovmar et al., 2003).

Table 4.1: Taxon Sampling

Family	Genus	Species name	Sex	Tissue source
Pitheciidae	*Pithecia*	*Pithecia irrorata*	M	BCNP
	Callicebus	*Callicebus donacophilus*	M	CRES
		Callicebus nigrifrons	M	CPRJ
	Cacajao	*Cacajao calvus*	F	CPRJ
		Chiropotes utahicki	M	BCNP
Atelidae	*Alouatta*	*Alouatta belzebul*	F	BCNP
		Alouatta caraya	M	BCNP
		Alouatta palliata	F	CRES
	Ateles	*Ateles belzebuth*	M	BCNP
		Ateles geoffroyi	F	UPM
		Ateles paniscus	M	Unknown
	Lagothrix	*Lagothrix lagotricha*	M	BCNP
	Brachyteles	*Brachyteles arachnoides*	M	CPRJ
Cebidae	*Callithrix*	*Callithrix argentata*	M	Unknown
		Callithrix flaviceps	M	CPRJ
		Callithrix geoffroyi	M	CPRJ
		Callithrix jacchus	F	WUSTL, BCM
		Callithrix kuhlii	M	CPRJ
		Callithrix penicillata	M	BCNP
	Callimico	*Callimico goeldii*	M	BCNP
	Leontopithecus	*Leontopithecus chrysomelas*	M	CPRJ
	Saguinus	*Saguinus fuscicollis*	M	BCNP
		Saguinus imperator	M	BCNP
		Saguinus martinsi	M	CPRJ
		Saguinus niger	M	BCNP
	Aotus	*Aotus azarae*	M	BCNP
		Aotus lemurinus	F	CBS
		Aotus nancymaae	M	UTMD
		Aotus vociferans	M	UTMD
	Cebus	*Cebus albifrons*	M	BCNP
		Cebus apella	M	BCNP
		Cebus nigritus	M	CPRJ
		Cebus olivaceus	F	BCNP
		Cebus xanthosternos	M	CPRJ
	Saimiri	*Saimiri sciureus*	M	BCNP
		Saimiri ustus	M	CBS
		Homo sapiens	M/F	GRC
		Macaca mulatta	F	MGSCBC
		Pan troglogytes	M	WUGSC, BI
		Pongo pygmaeus	F	WUGSC

BCM- Baylor College of Medicine Human Genome Sequencing Center, BCNP- Brazil Centro Nacional de Primates, BI- Broad Institute, CBS- Dr. Caro-Beth Stewart, CPRJ- Centro de Primatologia de Rio de Janeiro, CRES - The Center for Reproduction of Endangered species at the San Diego zoo, GRC- Genome Reference Consortium, MGSCBC- Macaque Genome Sequencing Consortium, Baylor College, UPM- The University of Pittsburg Medical Center, UTMD- The University of Texas MD Anderson Cancer Center, WUGSC- Washington University Genome Sequencing Center

Molecular Markers

Development of non-genic markers from shotgun libraries

Sequences obtained from genomic shotgun libraries and cloned in three representative platyrrhine species (Jameson et al., 2012) were parsed for non-genic marker sequences. Non-genic sequences were defined as those sequences that 1) do not reside in gene regions (exon, intron, UTR), 2) do not contain any repetitive elements, and 3) maintain a moderate GC content. Further details regarding the process of identifying these non-genic markers is detailed in Wildman *et al.* (2009).

Amplification and Alignment of non-genic markers

Primers for non-genic markers were designed by eye, within areas of conservation, using multiple sequence alignments of the shotgun sequences and the available genomes of human, chimpanzee, orangutan, macaque, and marmoset. Sequences from publicly available genomes were collected by carrying out BLAT searches in the UCSC genome browser (http://genome.ucsc.edu/) on the following genome builds: hg19, panTro3, ponAbe2, rheMac3, and calJac3. Non-genic markers were amplified using the primers presented in Table S6. PCR products were visualized on agarose gels and purified using the Qiagen MinElute 96 PCR purification kit (Qiagen, Valencia CA.). Purified PCR products were sequenced in both directions using Big Dye v.3.1 (Applied Biosystems, Foster City, CA). The sequences for each of the 64 amplified non-genic markers as well as those from species with publicly available genomes were individually aligned using the ClustalW algorithm (Thompson et al., 1994), examined by eye, and manually adjusted if necessary. All amplified sequences have been deposited into GenBank, accession numbers: KC760209-KC762207.

Phylogenetic analysis

Maximum parsimony analysis

Maximum parsimony (MP) analyses were carried out in PAUP* v.4.0b10 (Swofford, 2002) on a concatenated data matrix that included all markers. MP analyses consisted of initial heuristic searches of 1000 random addition sequence replicates with TBR branch swapping. MP bootstrap analyses of 1000 replicates, with 100 random addition sequence replicates per bootstrap, were used to obtain measures of branch support.

Maximum likelihood analysis

Maximum likelihood (ML) analysis was conducted in RAxML-7.3.2 and implemented through the CIPRES Science Gateway (Miller et al., 2010; Stamatakis, 2006). A data partitioning scheme of 64 partitions, representing the individual markers, was chosen as the best scheme as detailed below. The best-known likelihood tree was inferred from the partitioned dataset, after conducting 1000 bootstrap replicates with a random starting tree. Shimodaira-Hasegawa (SH) topology tests (Shimodaira and Hasegawa, 1999) were conducted in PAUP*. This test compares the likelihood score of trees with different topologies to determine if any topology is significantly more likely than the others given the sequence data.

Bayesian inference analysis

The most likely model of nucleotide substitution was determined using MrModeltest (Nylander, 2004) for the concatenated dataset as well as for each of the individual markers. To determine the most appropriate partitioning scheme for the data the harmonic means of the Bayesian analysis were used to calculate the Bayes factor

(B_{01} =Harmonic mean L_0/Harmonic mean L_1). The criterion of $2\ln B_{01} \geq 10$ was used to assess whether to accept the more parameterized model (Brandley et al., 2005; Kass and Raftery, 1995; Newton et al., 1994; Suchard et al., 2001). Upon comparing partitioning schemes with 1) a single partition, 2) 64 individual marker partitions, or 3) partitions based on GC content, we identified the most parameterized model (2) as the most appropriate partitioning scheme (Table S7).

A Bayesian inference (BI) phylogenetic analysis was conducted using MrBayes 3.2.1 implemented through the CIPRES Science Gateway (Miller et al., 2010; Ronquist and Huelsenbeck, 2003). The analysis was conducted on the 64 partition dataset with the models of nucleotide substitution chosen for the individual partitions. The analysis consisted of 2 runs with 4 chains each, sampled every 1000 generations, and run for 22 million MCMC generations, with a burn-in of 25% as determined by convergence statistics. In all phylogenetic analyses human, chimpanzee, orangutan, and macaque were used as outgroup taxa to root the tree.

Divergence date estimates

Divergence date estimates were calculated from the non-partitioned 40 taxa x 40,986 bp data matrix. The analysis was conducted using BEAST 1.7.2 with the BEAGLE framework implemented through the CIPRES Science Gateway (Drummond et al., 2012; Miller et al., 2010). Dates were estimated using a relaxed uncorrelated lognormal molecular clock that allows each branch its own rate of evolution, estimated independently from the neighboring branches, creating a lognormal distribution of the rates. A Yule process speciation tree prior was implemented and provides uniform speciation and an equal chance for speciation at all tips of the tree. The dating analysis

was carried out under a general time reversible (GTR) model that assumes a symmetrical substitution matrix, variable nucleotide rate frequencies and occurrence frequencies. Four rate categories were included and rate frequencies were set to a gamma distribution (γ). Minimum age calibration points for the tree are as follows: the most recent common ancestor (mrca) of *Homo* and *Pan* was estimated to be 5.7 million years ago (mya) based on the fossil hominin *Sahelanthropus* (Lebatard et al., 2008), the mrca of crown Catarrhini was estimated to 20.5 mya based on the fossil ape *Morotopithecus* (Gebo et al., 1997), the mrca of crown Pitheciidae was estimated to 15.7 mya based on the inclusion of the fossil platyrrhine *Proteropithecia* (Kay et al., 1998) and the mrca of *Cebus* and *Saimiri* was estimated to be 12.2 mya based on *Neosaimiri*, an extinct cebid genus closely affiliated with *Saimiri* (Takai, 1994). An additional analysis was carried out with parameters identical to those described above with the addition of another age calibration point. In the second analysis, the age of crown Platyrrhini was estimated at 25.7 mya, assuming the fossil *Branisella* (MacFadden, 1990), the earliest known Neotropical primate, was a crown platyrrhine. Both analyses were independently run for 80 million generations and sampled every 1,000 generations with a burn-in of 50% as determined by standard convergence statistics (Ronquist and Huelsenbeck, 2003).

Biogeographic analysis

Biogeographic ranges

The geographic ranges of platyrrhine species were based on species distribution data obtained from both BDGEOPRIM and the IUCN redlist (Hirsch et al., In Prep.; IUCN, 2012). These species ranges were next included in one or more of nine total

operational geographic units (OGU) spanning South and Central America. OGUs were based on interfluvial regions simplified from those defined by Goldani *et al.* (2006) as seen in Figure 4.1. These areas include: (A), Central America and Mexico bounded on the south by the Magdalena river; (B), Venezuela, Guyana, Surinam, French Guiana and portions of Brazil and Colombia bounded by the Magdalena river on the west, the Negro and Amazon rivers on the south and the Caribbean coast; (C), Ecuador and portions of Colombia, Peru and Brazil bounded by the Negro river on the north, the Maranon and Amazon rivers on the south, and the Pacific coast on the west; (D), Peru bounded by the Ucayali river on the east, the Maranon river on the northwest and the Pacific coast; (E), portions of Peru and Brazil bordered by the Amazon river on the north, the Ucayali river on the west, the Tapajos river on the east and the Madre de Dios and Guapore rivers on the south; (F), Brazil bordered by the Atlantic coast and the Amazon river on the north, the Tapajos and the Teles Pires rivers on the west and the Araguaia river on the east; (G), Brazil and Uruguay bordered by the Atlantic coast on the north and east and the Araguaia and Parana rivers on the west; (H), Brazil bordered by the Teles Pires river on the east and the Tapajos river on the west; and (I), Argentina, Chile, Bolivia, Paraguay and a portion of Brazil bordered by the Madre de Dios and Guapore rivers on the north, the Parana river on the east and the coastline on the south and west. Non-platyrrhine outgroup species (human, chimpanzee, orangutan, and macaque) were coded as residing in a 10th OGU (J) representing all geographic space outside the 9 above specified OGUs.

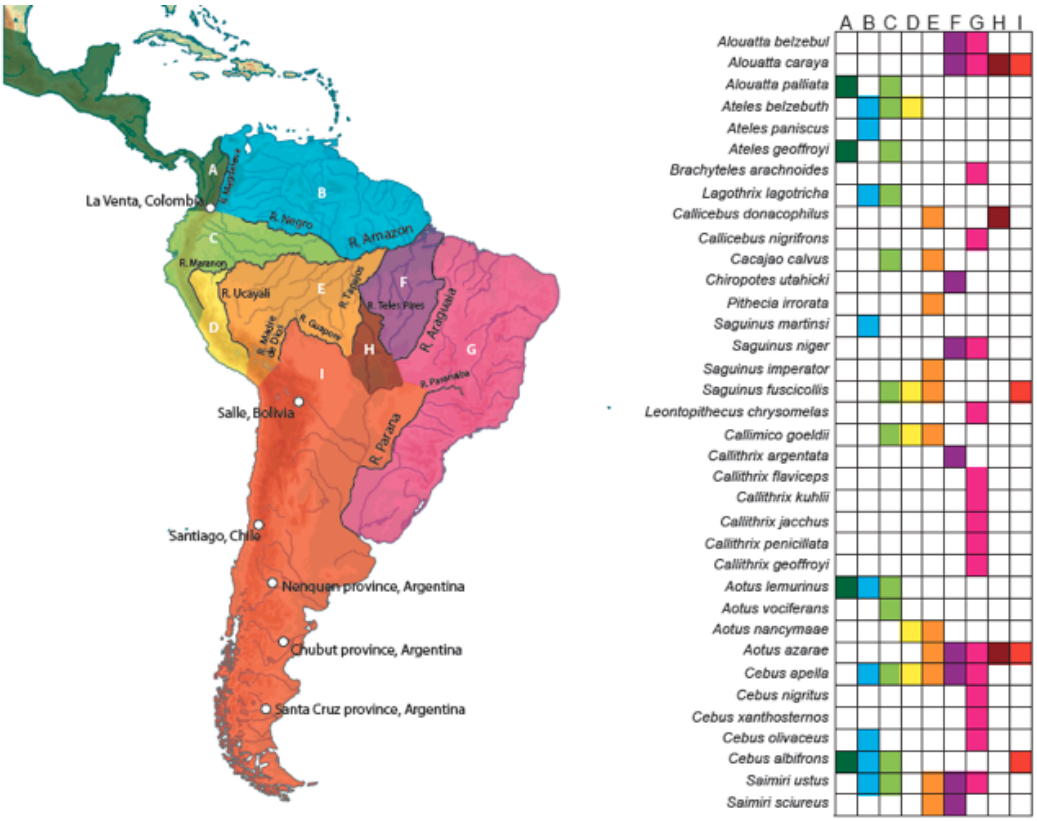


Figure 4.1: Geographic ranges and distributions of New World monkeys.

Map of the Neotropics showing nine geographic regions and the rivers that define them and the distribution of the 36 NWM species among these ranges determined using distribution data obtained from BDGEOPRIM and IUCN redlist. Colored boxes indicate presence of species in the region.

Ancestral area reconstruction

Ancestral area reconstructions of platyrrhines were carried out using the RASP package (Reconstruct Ancestral State in Phylogenies, version 2.0b) (Yu et al., 2010; Yu et al., 2011). The RASP package includes S-DIVA, a statistical dispersal-vicariance analysis program that predicts ancestral ranges by giving extinction and dispersal events a greater cost than vicariance events while taking into account phylogenetic uncertainty. In addition, RASP also includes a modified Bayesian binary MCMC methodology (BBM) to predict ancestral areas through multiple MCMC iterations. To account for uncertainties in the platyrrhine phylogeny, S-DIVA and BBM were run using 10,000 randomly chosen post-burn-in trees from the BEAST MCMC output. The number of maximum areas in the S-DIVA analysis was limited to two. Ten MCMC chains were

ran simultaneously for 5,000,000 generations with a sampling every 100 generations. The root distribution was set to outgroup and the analysis was run under the Estimated F81+G (Felsenstein 81+ γ) Model. The number of maximum areas was set also two.

Results

Non-genic molecular markers

A total of 64 molecular markers served as the basis for this study (GenBank, accession numbers: KC760209-KC762207). Markers ranged in length from 241 bp to 1258 bp with a mean length of 629 bp (SD = 215 bp) and a total concatenated length of 40,986 bp. The GC content of the markers ranged from 26% to 58% with a mean of 38% (SD = 6.59%). BLAT searches to the human genome show that each of the human chromosomes contains at least one of the 64 markers with the exception of chromosomes 22 and Y. The complete concatenated data matrix consisted of 40 taxa x 40,986 bp. Details on the annotation of each marker (length, GC content, genome location, and taxon coverage) are presented in Table S8.

Phylogenetic analysis

MP, ML, and BI analyses were conducted on the 40 taxa x 40,986 bp data matrix. All analyses converged on a single optimal branching arrangement among species (Figure 4.2). Bootstrap scores (BS) and posterior probabilities for nodes with less than full support (MP< 100, ML< 100, BI<1.0) are described below and are shown in Figure 4.2.

Maximum parsimony analysis

A single optimal MP tree was inferred. This tree has 7,518 parsimony informative characters and a minimum length of 17,951. There are four nodes that did not receive

full support: *Aotus*+callitrichines (BS = 93), *Cebus olivaceus*+*Cebus albifrons* (BS = 69), *Callithrix penicillata*+*Callithrix geoffroyi* (BS = 62), and *Callithrix jacchus*+(*Callithrix penicillata*/*Callithrix geoffroyi*) (BS = 92).

Maximum likelihood analysis

The ML analysis was carried out on a dataset separated into 64 individual non-genic marker partitions. The final tree topology has an optimized likelihood of ln *L* -166,811.88. There are three nodes that did not receive full support: *Aotus*+callitrichines (BS = 92), *Cebus olivaceus*+*Cebus albifrons* (BS = 87), and *Callithrix penicillata*+*Callithrix geoffroyi* (BS = 80).

Bayesian inference analysis

Bayesian inference was carried out on the dataset separated into 64 individual non-genic marker partitions. The single optimal tree topology had an optimized likelihood of ln *L* -166,811.88. There are two nodes that did not receive full support: *Cebus olivaceus*+*Cebus albifrons* (BI = 0.99), and *Callithrix penicillata*+*Callithrix geoffroyi* (BI = 0.99).

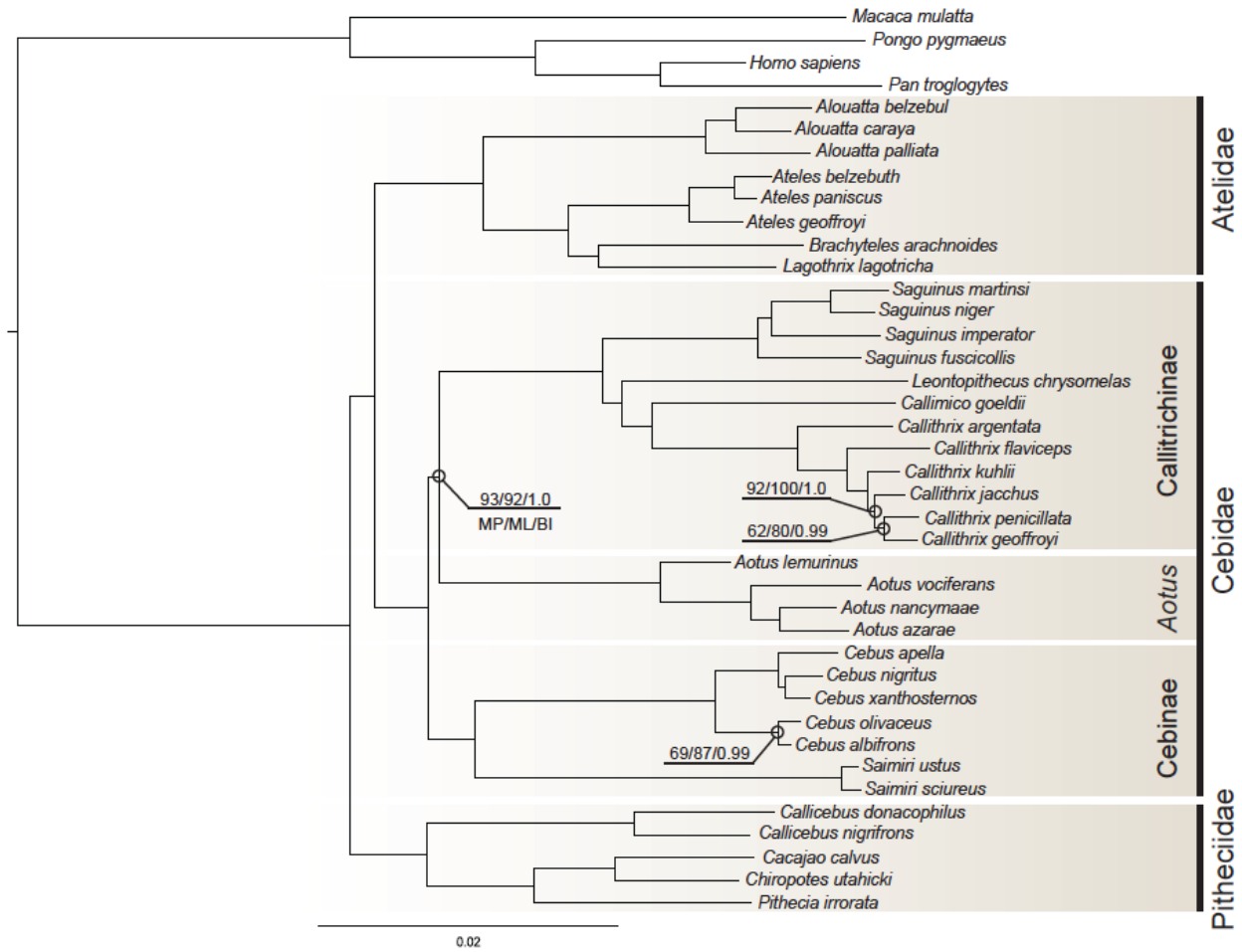


Figure 4.2: Phylogeny of extant NWM species.

The consensus tree according to maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI) analysis of the 40,986 bp concatenated multiple sequence alignment of 64 non-genic markers. Branch lengths shown are from the ML analysis using RAxML. Bootstrap and posterior probability scores are shown on nodes that did not receive full support (MP/ML/BI = 100/100/1.0).

Partitioning scheme and model selection

To determine the partitioning scheme to best represent our dataset, three different partitioning schemes were tested using a Bayes factor analysis. The three partitioning schemes include 1) a single concatenated partition, 2) GC content partitions, and 3) 64 individual marker partitions. Scheme 2 contains a total of 6 partitions based on the sequence's GC content. The markers were plotted by GC content and placed into bins representing the lowest, low to mid, mid to high and

highest GC contents. In addition we removed the upper and lower outliers into their own bins, for a total of six distinct partitions. See Table S7 for more information on the partitions.

Prior to the Bayes factor test it was necessary to determine the most appropriate model of sequence evolution for each of the three above partitioning schemes. Results from MrModeltest based on the AIC found that GTR+γ was the most appropriate model of sequence evolution for scheme one. GTR + γ was chosen as the best fit model for all partitions in scheme two, except for the upper outliers, for which SMY + γ was chosen as the best fit model. For the majority of partitions in scheme 3 GTR + γ was chosen as the best-fit model (see Table S7 for full details). Bayes factor analysis found that the most partitioned scheme (scheme three, 64 individual markers) was preferred over the other partitioning schemes and the most appropriate to use in further Bayesian analysis.

Hypothesis testing

In order to test whether the optimal branching arrangement as established by phylogenetic analyses is significantly better than alternate arrangements, a Shimodaira-Hasegawa (SH) topology test (Shimodaira and Hasegawa, 1999) was carried out. The likelihood scores of the optimal branching arrangement joining *Aotus*, Cebinae and Callitrichinae along with the likelihood scores of the other alternate arrangements of the three branches were compared. The optimal topology joining *Aotus* and the callitrichines to the exclusion of the Cebinae (as shown in Figure 4.2) had the highest likelihood score of ln *L* -173,521.32, followed by the joining of Cebinae with Callitrichinae (ln *L* -173,542.67) and lastly the joining of *Aotus* with Cebinae (ln *L* -

173,542.42). However the optimal topology was not found to be statistically significantly better than the alternate topologies (p-values = 0.1061 and 0.1067 respectively).

Divergence dating

The BEAST divergence date estimates among the 40 species used in the phylogenetic analysis are shown in Figure 4.3 and Table 4.2. Divergence date estimations were carried out using two separate schemes of fossil calibration schemes. The analysis that did not include *Branisella* as a calibration point estimated the most recent common ancestor (mrca) of all platyrrhines and catarrhines at 30.69 mya, with a 95% credibility interval between 23.33 and 56.73 mya. The mrca of extant platyrrhines is estimated at 18.27 mya with a 95% credibility interval between 13.04 and 40.16 mya. Among the platyrrhine families the mrca of cebids is estimated at 14.99 mya, the mrca of pitheciids at 14.21 mya, and the mrca of atelids at 12.27 mya.

The divergence date analysis including *Branisella* as a fossil calibration point, representing the minimum age of crown platyrrhini, estimated divergence dates that are overall older than found in the previous analysis. The mrca of all platyrrhines and catarrhines is estimated at 39.05 mya, with a 95% credibility interval between 29.63 and 49.27 mya. The mrca of extant platyrrhines is estimated at 25.32 mya, with a 95% credibility interval between 25.06 and 28.19 mya. Among the platyrrhine families, the mrca of cebids is estimated at 20.75 mya, the mrca of pitheciids at 17.79 mya, and the mrca of atelids at 16.64 mya.

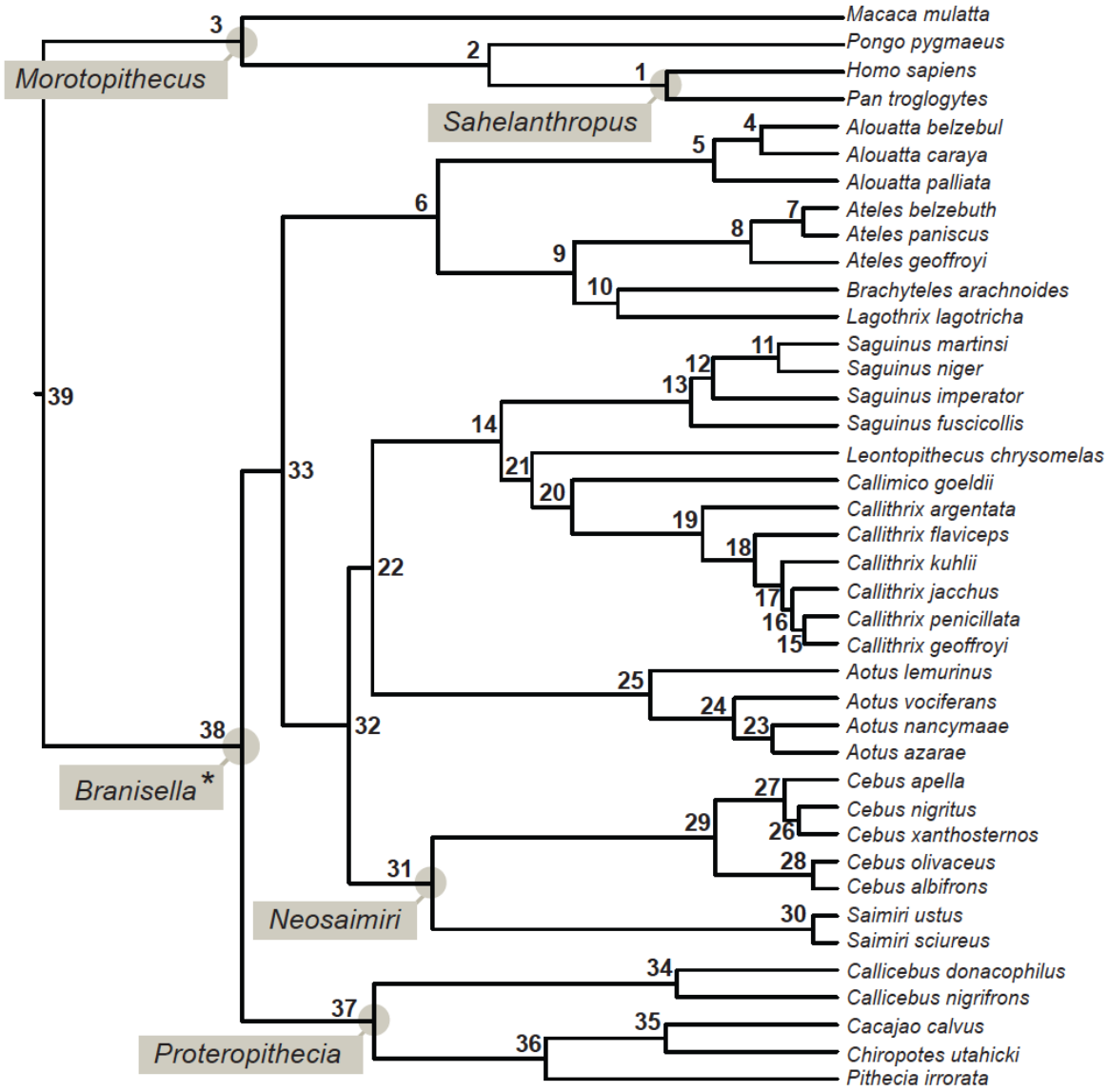


Figure 4.3: Divergence date analysis.

NWM divergence date tree as estimated using BEAST. Shaded circles and boxes signify nodes and fossils that were used as calibration points. Two independent analyses were conducted, one using fossil calibrations including Morotopithecus, Sahelanthropus, Neosaimiri and Proteropithecia the other including the above and also Branisella. Table 4.2 show the median estimates in million years ago (mya) for the age of each node as well as the 95% highest posterior density interval (HPD) for both analyses.

Table 4.2 Divergence dates at nodes

Excluding *Branisella*			Including *Branisella*	
Median Age (MYA)	95% HPD	Node	Median Age (MYA)	95% HPD
5.43	3.98 - 7.73	**1**	6.22	3.80 - 12.27
10.88	6.72 - 17.36	**2**	13.21	9.29 - 17.86
18.47	17.07 - 22.81	**3**	21.23	17.21 - 27.66
2.33	0.03 - 3.59	**4**	3.23	0.29 - 4.91
3.77	0.17 - 8.51	**5**	5.06	0.91 - 7.38
12.27	6.17 - 21.78	**6**	16.64	9.26 - 21.57
1.03	0.12 - 3.03	**7**	1.37	0.19 - 2.21
2.64	0.24 - 5.75	**8**	3.62	0.32 - 5.30
8.07	3.42 - 13.65	**9**	10.96	3.37 - 14.32
6.73	2.67 - 10.24	**10**	9.11	2.58 - 12.79
1.80	0.03 - 6.92	**11**	2.38	0.06 - 3.56
3.80	0.35 - 5.80	**12**	5.18	0.12 - 6.84
4.48	0.44 - 9.44	**13**	6.02	0.37 - 7.90
10.32	7.40 - 18.99	**14**	13.83	5.17 - 16.86
1.00	0.04 - 1.59	**15**	1.39	0.14 - 2.00
1.37	0.13 - 3.13	**16**	1.84	0.32 - 2.59
1.68	0.26 - 4.92	**17**	2.25	0.61 - 3.19
2.52	0.35 - 9.32	**18**	3.37	1.30 - 4.70
4.13	1.65 - 7.02	**19**	5.60	1.55 - 7.47
8.13	5.61 - 11.56	**20**	11.26	3.17 - 14.09
9.37	6.66 - 13.43	**21**	12.90	4.70 - 15.86
14.28	2.97 - 24.67	**22**	20.10	7.38 - 22.94
1.98	0.06 - 3.03	**23**	2.74	0.20 - 3.96
3.16	1.12 - 5.76	**24**	4.19	0.98 - 5.79
5.75	3.24 - 19.95	**25**	7.57	2.21 - 10.37
1.18	0.09 - 2.08	**26**	1.69	0.04 - 2.93
1.61	0.35 - 4.76	**27**	2.31	1.02 - 9.35
0.74	0.02 - 5.32	**28**	0.94	0.05 - 3.80
3.74	1.13 - 7.41	**29**	5.26	3.27 - 8.80
0.73	0.02 - 7.62	**30**	0.97	0.12 - 2.01
12.42	10.42 - 15.03	**31**	17.15	10.68 - 20.10
14.99	11.82 - 27.34	**32**	20.75	15.37 - 23.66
17.02	13.27 - 28.59	**33**	24.17	19.84 - 26.62
4.92	0.90 - 12.71	**34**	6.34	1.67 - 9.38
5.26	2.52 - 8.89	**35**	6.45	0.52 - 9.07
8.95	4.04 - 13.32	**36**	11.25	4.17 - 15.53
14.21	10.97 - 18.34	**37**	17.79	12.74 - 21.61
18.27	13.04 - 40.16	**38**	25.82	25.07 - 28.20
30.69	23.34 - 56.74	**39**	39.04	29.64 - 49.27

Rate variation among clades

The rates of sequence evolution on all branches were determined by the Bayesian divergence estimate analysis. A Mann-Whitney test determined that there is a significant difference between the rates within pitheciids when compared to the remaining NWMs as well as within callitrichines when compared to the remaining NWMs (Table 4.3). The pitheciids were found to have a mean rate of 0.008 substitutions/site/ million years, lower than the non-pitheciid NWMs (0.022, $p<0.005$, Mann-Whitney). Conversely, callitrichines were found to have a mean rate of 0.024, significantly greater than the non-callitrichine NWMs (0.017, $p<0.005$, Mann-Whitney).

Table 4.3 Rate variation

Taxa groups	Mean rate (sub/site/ Million years)	p value (Mann-Whitney)
Pitheciids	0.008	< 0.005
Non-pitheciids	0.022	
Callitrichines	0.024	< 0.005
Non-callitrichines	0.017	

Reconstruction of platyrrhine biogeographic history

Ancestral area reconstructions were conducted in RASP and carried out using Dispersal-Vicariance (S-DIVA) and Bayesian Binary MCMC (BBM) methodology (Yu et al., 2010; Yu et al., 2011). A graphical representation of the results from these analyses is shown mapped onto a tree in Figure 4.4. S-DIVA results are displayed in the black outlined circles, and BBM results are displayed in the grey outlined circles at each node. The colors in the pie charts represent the ranges, and the size of the pie pieces represents the RASP value, i.e. the frequency of that range occurring at that node.

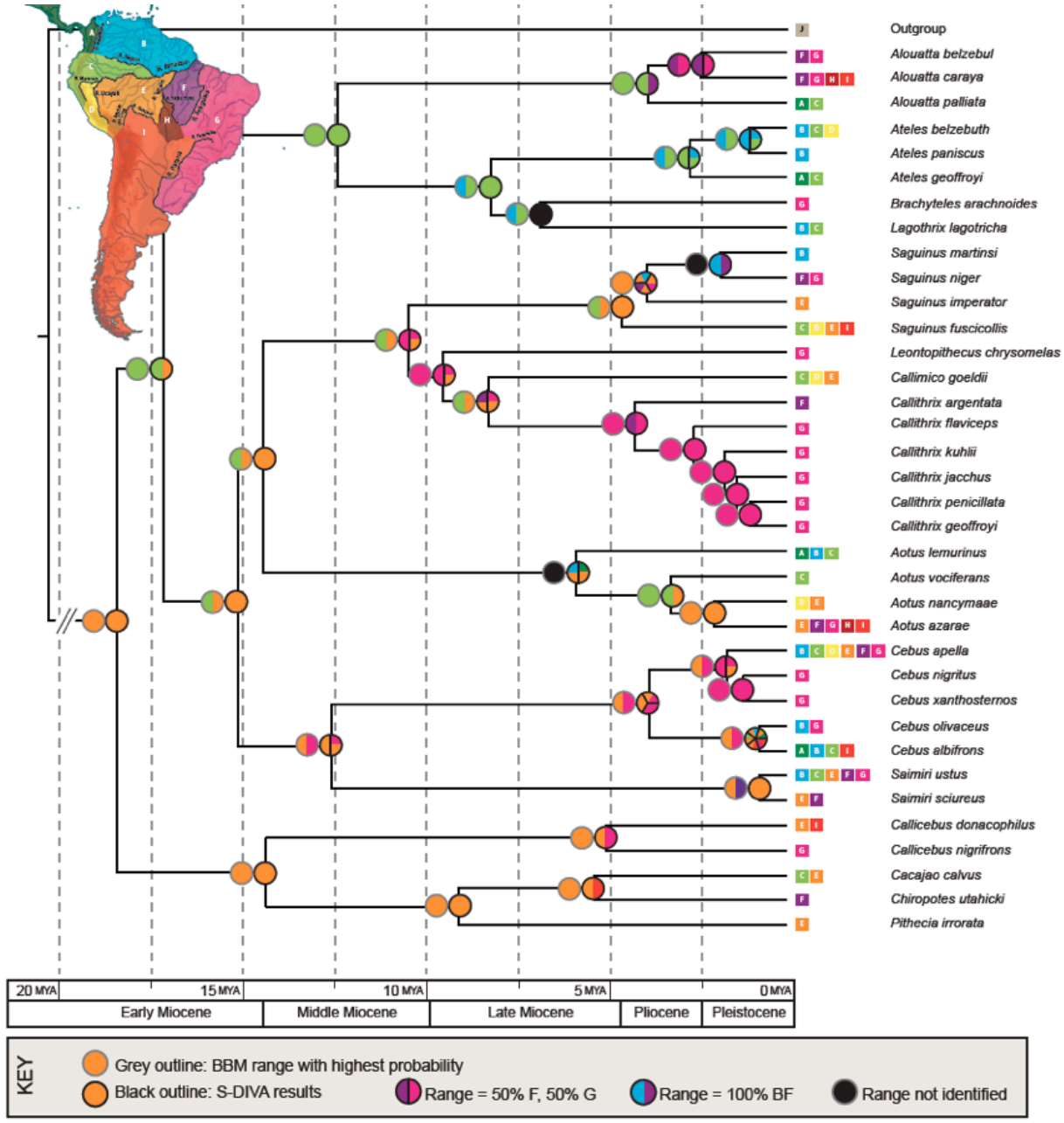


Figure 4.4: Ancestral area reconstruction.

Results of ancestral area reconstruction from S-DIVA and BBM analysis overlaid on the chromatogram from the divergence date analysis. Present day distributions of NWM taxa are shown on the branch tips as colored boxes. The non-platyrrhine outgroup species have been collapsed into a single branch for simplicity. Colors within pie charts signify the ranges and the size of the wedges show the relative likelihood of alternate scenarios. Grey outlined pie charts at nodes show the range found to have the highest probability as determined by the BBM analysis. Black outlined pie charts at nodes show the ancestral distribution inferred by S-DIVA. Range colors separated by a black line indicate separate ranges, colors without separation indicate vicariance events involving the regions indicated by those colors (see key in figure). Nodes with black circles indicate positions where no reconstruction could be inferred.

Black lines separating the pie pieces indicate independent range predictions, pie pieces without black boarders represent shared range predictions. Both methods of ancestral area reconstruction produced similar results at the majority of nodes in the tree.

A summary of the key results from the biogeographic analyses follows and is depicted in Figure 4.5. We found that the mrca of extant platyrrhines likely resided in western South America, with populations found specifically in portions of Peru and Brazil (E). This prediction is supported with a RASP score (RS) of 100% from both S-DIVA and BBM methods. The divergence of the ancestors of present day pitheciids from other NWMs began in the early Miocene, during which time early NWMs began a northern expansion in which they also inhabited Ecuador and portions of Colombia, Peru and Brazil (C). BBM analyses show that the ancestral population was living solely in the northern region (C, RS = 100) while S-DIVA analyses predict the ancestral population to be split between the two regions (E and C, RS = 100). This persisted until the end of the early Miocene when the lineage leading to extant Atelidae diverged and migrated out of the southern region (E) and into (C). Extant Atelidae genera, including *Alouatta*, *Ateles*, *Brachyteles* and *Lagothrix*, today reside mainly in coastal regions.

As recently as the middle Miocene the descendants of the mrca of cebines, a clade that today contains *Cebus* and *Saimiri*, began to expand geographically. Both S-DIVA and BBM analyses agree that some of their ancestors resided in portions of Peru and Brazil (E). However the methods disagree on additional regions of ancestral areas. S-DIVA analysis predicts with a frequency of 50% (RS = 50) that the ancestors lived solely in region E, or also with a frequency of 50% (RS = 50) that the ancestors inhabited both region E and more temperate areas of southern South America (I). BBM

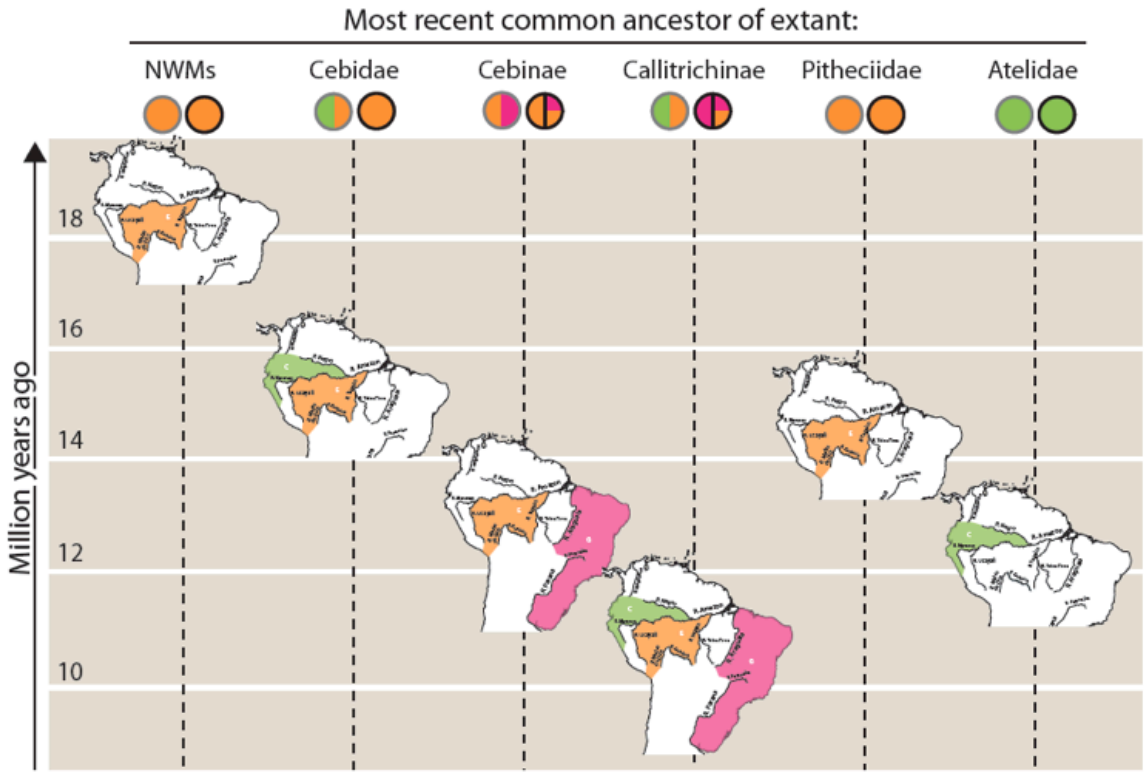


Figure 4.5: Depiction of clade distribution throughout South America.

The predicted ranges of the mrca of extant platyrrhine clades is shown on a map of present day South America along a time scale (mya) detailing the estimated times of clade divergence. Circles show the ancestral area reconstruction results (shown for all nodes in Figure 4.4). As in Figure 4.4, colors within pie charts signify the ranges and the size of the wedges show the relative likelihood of alternate scenarios. Grey outlined pie charts at nodes show the range found to have the highest probability as determined by the BBM analysis. Black outlined pie charts at nodes show the ancestral distribution inferred by S-DIVA. Range colors separated by a black line indicate separate ranges, colors without separation indicate vicariance events involving the regions indicated by those colors (see key in figure 4.4).

analyses predict (RS = 100) the early cebines lived in both region E and Brazil and Uruguay bordered by the Atlantic coast on the north and east and the Araguaia and Parana on the west (G). Descendants of these ancestors now reside in a variety of regions, with *Cebus* and *Saimiri* species having some of the most widespread ranges (e.g. the present day regions included in the ranges of *Cebus apella* B, C, D, E, F, and G and *Saimiri ustus* B, C, E, F, and G).

Beginning in the late Miocene the earliest callitrichine diverged from other cebids. Again, in some instances, the ancestral area reconstructions differ based on the two methods used. S-DIVA predicted, with a frequency of 50% (RS = 50), that the ancestral

callitrichines lived in Brazil and Uruguay bordered by the Atlantic coast on the north and east and the Araguaia and Parana on the west (G). Also, with a frequency of 50% (RS = 50), these ancestors were predicted to have resided in both region G and portions of Peru and Brazil (E). This is in contrast to the BBM analysis (RS = 100) that estimated the ancestral range of the mrca of callitrichines to be shared among region E and Ecuador and portions of Colombia, Peru and Brazil (C). Today the callitrichines have slightly varied ranges however the marmosets (genus: *Callithrix*) reside solely in regions E and F. Ancestral areas predictions for all ancestral nodes, as well as RASP scores, are shown in Figure 4.4.

Discussion

Phylogenetic signal of non-genic markers

Here we have shown that the use of non-genic markers in a large-scale (>40 kb) phylogenetic analysis was able to resolve, with high confidence, the branching arrangements within the New World monkey clade. Many phylogenetic studies have described the importance using neutrally evolving sequence markers for phylogenetic reconstruction (Peng et al., 2009; Siepel et al., 2005; Visel et al., 2008). To this end, non-genic markers have shown to be a powerful tool in the resolution of difficult to determine primate relationships that have been difficult to determine (Wang et al., 2012; Wildman et al., 2009). The power of non-genic markers is derived from their nature as neutrally evolving genomic DNA, which, in theory, is free of the constraints of selection pressure. The use of this kind of sequence data allows us to potentially remove much of the 'background noise' present within the sequences due to selection and focus only on the phylogenetic signal. Our data set suggests that these non-genic sequences are in

fact under fewer constraints, as they have more sequence variation among them than the genomes do as a whole. The amount of genetic diversity between human and chimpanzee in our non-genic dataset is 1.9%, an increase from 1.23%, the amount of diversity found between the genomes as a whole (Consortium, 2005). This increased sequence variation is also shown specifically within the NWM clade, where we observe an increase in the phylogenetic signal (number of substitutions/site) when compared to a genic dataset. For example, on the branch leading to *Aotus* the non-genic dataset has twice as many substitutions (ML branch length = 7.0×10^{4}) as compared to a similarly sized dataset using genic (intron and exon) sequences (Perelman et al., 2011). The non-genic sequences used in these analyses show an increased rate of sequence evolution, and thus appear to be under less selection pressure and evolving more freely than genic regions of the genomes as a whole.

Overall non-genic markers were able to resolve the NWM branching order with 88% of nodes receiving full support (MP/ML/BI = 100/100/1.0). The success of this dataset was found to be dependent on both the size and the type of data used. We found that the non-genic data is more powerful than traditional (genic) markers. For example, the similarly sized dataset using traditional markers by Perelman *et al.* (2011) resolved only 65% of the nodes within the NWM clade with full support, whereas we resolved 88% of nodes with full support. It has been recognized that increasing the size of a dataset will increase its phylogenetic power. We also found that the larger size of this dataset is substantially helped in resolving the nodes, as a smaller (7.6 kb) dataset using non-genic markers from Wildman *et al.* (2009) was able to resolve only 66.6% of the nodes with full support. Here we demonstrate that both increasing the amount of

data as well as choosing the appropriate type of data is equally as valuable and is a critical step in meaningful phylogenetic analyses (Philippe et al., 2011). The difficulty in obtaining non-coding markers from species that do not have reference genome assemblies has hampered their use in large-scale studies. However accessibility to the once termed 'junk' DNA is steadily growing, due to advances in sequencing technology and the increasing availability of whole genomes, which significantly increases their ease of use in future studies.

Fully resolved tree topology

This study produced a fully resolved, highly supported phylogeny among 36 New World monkey species. The branching arrangement of the NWM families is in agreement with some previously published studies, and supports the sister grouping of Atelidae and Cebidae, to the exclusion of Pitheciidae. The arrangement of NWM genera has been well established in the literature, with the exception of the genus *Aotus* (Perez et al., 2012; Wildman et al., 2009). We show, with high branch support, that *Aotus* is the sister taxa to Callitrichinae. This placement is consistent with previously published studies using single gene datasets (Goodman et al., 1998; Harada et al., 1995; Porter et al., 1997a, 1999) as well as a large scale (~35 kb) concatenated dataset of multiple nuclear genes (Perelman et al., 2011). Although several studies have produced different branching arrangements (Barroso et al., 1997; Canavez et al., 1999; Opazo et al., 2006; Perez et al., 2012; Schneider et al., 2001; Schneider et al., 1996; Wildman et al., 2009), most have had weak branch support. Given, the scale of these datasets (<15 genes/markers) and the low branch supports, it appears that these alternative branching arrangements may be due to homoplasy, or ancient incomplete lineage sorting, among

the specific genes included in the datasets. Given the information we have today on the confounding effects of short branches it is not surprising that previous phylogenetic studies based on single or few genes were unable to come to a consensus on the placement of *Aotus*. However, we do not consider the position of *Aotus* to be completely resolved, and advocate for continued investigation into this long-standing question. Although branch support scores at this node were high (MP=93, ML=92, BI=1.0), these measurements can often be misleading. Bootstrap scores, in essence, measure if the dataset is adequately sized in order to find a well-supported solution, thus when used as a measure of branch support for large datasets, this value becomes essentially meaningless (Felsenstein, 2004). We find hypothesis testing to be a more accurate measure of confidence in branching arrangements. Although all phylogenetic analyses conducted agreed that the grouping of *Aotus* with Callitrichinae was the optimal topology, it was not significantly favored by topology testing. Shimidiro-Hasigawa topology tests carried out on this dataset failed to find any one branching arrangement to be significantly more optimal than the other proposed branching arrangements. Due to this finding, we do not accept the support from this or any other study for the sister relationship of *Aotus* and Callitrichinae as completely definitive. However, it is interesting to note that if this branching arrangement is ultimately validated by more data, then the small body masses shared between Owl monkeys (*Aotus*) and callitrichines likely evolved before they last shared a common ancestor in the Miocene. We suggest further phylogenetic studies including increased species sampling of these taxa, as well as investigation of alternative, non-dichotomous branching evolution

scenarios that occurred possibly as a result of hybridization or incomplete lineage sorting.

The branching arrangements at the species level within our dataset are generally in agreement with those seen in the literature, with a few exceptions. Within the genus *Ateles* we find *A. belzebuth* to be sister to *A. paniscus,* to the exclusion of *A. geoffroyi*. However previous studies have excluded *A. paniscus* from the other *Ateles* species (Fabre et al., 2009; Nieves et al., 2005; Perelman et al., 2011). This inconsistency suggests that denser taxon sampling may be required to resolve contentious relationships *Ateles*. Within the genus *Aotus* the phylogenetic relationships among the species are fairly ambiguous. The genus was divided based on pelage into two groups, the grey-necked owl monkeys who live north of the Amazon River and the red-necked owl monkeys living to the south (Hershkovitz, 1983). However, recent molecular studies do not support grey and red-necked Owl monkeys as two monophyletic clades (Menezes et al., 2010). Our data also do not support a monophyletic grey-necked clade, instead they support a sister grouping of *A. azarae* and *A. nancymaae* both members of the red-necked group, followed by the branching of *A. vociferans* and finally *A. lemurinus*, both members of the grey-necked group (Figure 4.6). This arrangement is not consistent with a study of mitochondrial genes within the *Aotus* genus, which finds (*A. nancymaae* (*A. vociferans* (*A. azarae*, *A. lemurinus*))) (Menezes et al., 2010). However, the branching arrangement in the present study is consistent with phylogenies based on karyotypic data within the clade (Galbreath, 1983). Much of the phylogenetic uncertainty seen within this clade has been suggested to be a result of frequent species

misidentification due to morphological similarity, resulting in potentially mislabeled sequences in the public databases (Menezes et al., 2010).

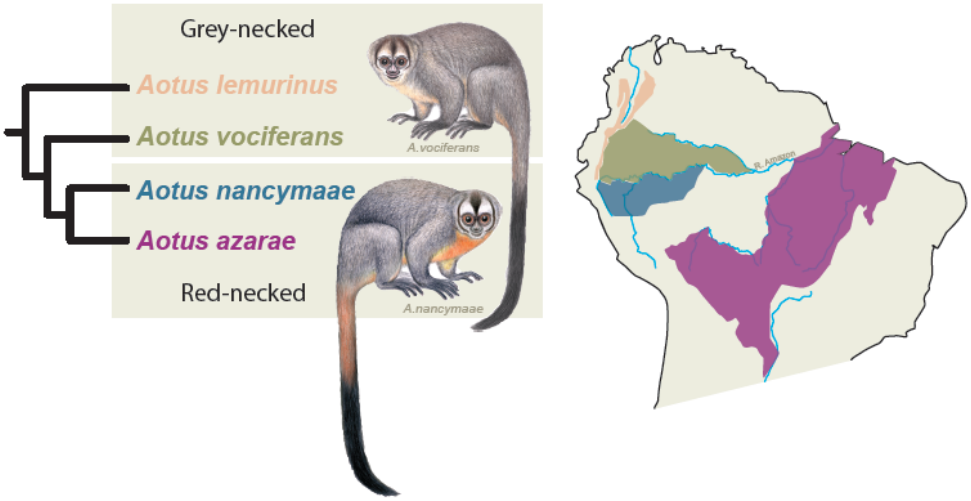


Figure 4.6: Phylogenetic resolution within Aotus.

Showing the phylogenetic relationship among species determined by our analysis and examples of grey and red-necked species. Monophyly of Grey-necked group is not retained. The map of South America shows the ranges (adapted from IUCN redlist) of *A. lemurinus, A. vociferans, A. nancymaae, and A. azarae*. Illustrations copyright 2012 Stephen D. Nash/IUCN/SSC Primate Specialist Group. Used with permission.

Variation in evolutionary rates

It has long been asserted that differences in the rates of nucleotide substitution are determined by variation in life history traits (Goodman, 1961, 1962; Li et al., 1985; Ohta, 1998). Within our data we find that callitrichines (marmosets, tamarins, and *Callimico*) have evolved at a faster pace than the other New World monkeys. This result is somewhat expected given that callitrichines have a faster life history than any other NWM group. Callitrichines are very small primates with a relatively short lifespan; they are also the only anthropoid primates to consistently produce litters of 2 or more offspring (Goldized, 1987). The average age at first reproduction in callitrichines is 21.5

months and the inter-birth interval is 7 months, both far less than the average among all platyrrhines (48.4 months and 17.6 months, respectively, Jones et al., 2009).

Conversely, when looking at the rates of nucleotide substitution within the pitheciids we see a slow down relative to the rest of the NWMs. Reductions in the rates of substitution in specific lineages have been seen in other primates, particularly in the human lineage when compared with the hominids, and in the chimpanzee lineage when compared to gorillas and orangutans (Bailey et al., 1991; Bailey et al., 1992; Elango et al., 2006; Goodman, 1961, 1962). These slow downs are often thought to be due to changes in life history traits, specifically those that extend generation time. The long interbirth interval seen in *Cacajao*, a pitheciid, of 25.83 months is substantially longer than the average, and out of the 95% confidence interval of the interbirth interval for all NWMs (mean = 17.6 months, Jones et al., 2009). The expansion of an interbirth interval could play an important role in lengthening the generation time of an organism, thus resulting in a decreased rate of evolution.

Primate dispersal to South America

South America was an island continent for at least 60 million years from the time Gondwana broke apart in the late Cretaceous until joining with Central/North America in the Pliocene (Ciochon and Chiarelli, 1980a). Several organisms such as pantodonts, uintatheres and xenungulates inhabited South America at the time of the split from Africa (Janis, 1998), others including cricetid rodents, procyonid and mustelid carnivores, and tayassuid artiodactyls migrated to the continent after the emergence of the Panamanian land bridge, during the Great American Biotic Interchange (Webb, 1991). Remarkably others, including primates and caviomorph rodents, journeyed to

South America during its period of isolation between these events. There is a large body of research centered on the mode in which primates arrived in South America that has developed three most probable scenarios; 1) migration through North America to South America, 2) migration through Antarctica to South America, and 3) migration directly to South America (Figure 4.7).

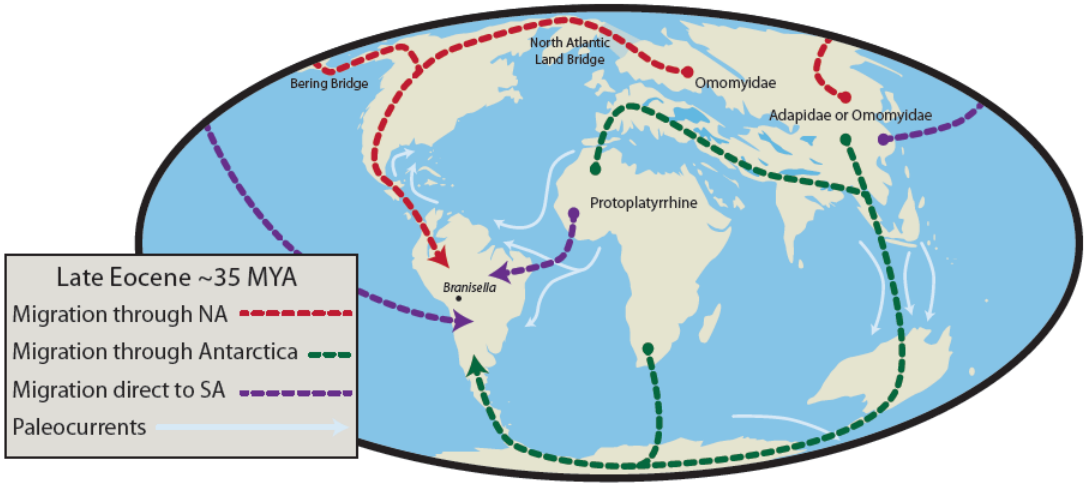


Figure 4.7: Hypothesized migratory routes to the Neotropics.

Map showing proposed routes of migration of primates to South America and their intermediate steps along the way. Red lines show routed leading to SA through NA, green lines show routes leading to SA through Antarctica, and purple lines show routes leading directly to SA from the ancestral primates. Light blue lines show predicted paleocurrents during the late Eocene.

A platyrrhine migration through North America (NA) was first formally proposed by Simpson (1945) and supported by the presence of faunal interchange between North and South America. This scenario proposes Eocene Omomyidae, living in continuity in both NA and Europe, and inferred as the ancestral anthropoid primate. These populations of omomyids maintained free gene flow, until the beginning of the middle Eocene when the North Atlantic Bridge became submerged, effectively splitting the putative ancestors of platyrrhines and catarrhines. The platyrrhines are then thought to have continued south through the Caribbean Sea to South America (SA) via volcanic

island arcs and waif dispersal. Waif dispersal, or island hopping, refers to the geographic spread of an organism across a barrier with which it has low probability of being able to cross (Simpson, 1978). In this case island hopping refers to travel along a string of islands, through the Caribbean to SA, likely drifting between islands on vegetation rafts. This mode of migration has received support largely on geophysical grounds, as the sea barrier between NA and SA was smaller relative to that between SA and Africa and the Caribbean island arc system was available as a passive transport mechanism (Cachel, 1981). Another variation of this model consists of ancestral platyrrhines arriving in NA from an Asian anthropoid ancestor and migrating to SA. This scenario has been based on either an adapid (Gingerich, 1980) or omomyid (Delson and Rosenberger, 1980) ancestor of proto-platyrrhines who dispersed from North America to South America (Ciochon and Chiarelli, 1980b). However, within any North American migration model for mammals the weaknesses come in crossing the Caribbean (Figure 4.7). Consideration of the ocean currents in the Caribbean make travel from NA to SA less likely because they flowed east to west and south to north, thus carrying any floating object from South America north to Central America, not vice versa (Berggren and Hollister, 1974).

There has also been support for scenarios by which protoplatyrrhines reached South America via Antarctica and the Gondwanan continents. Prior to the opening of the Drake Passage in the late Eocene (~37 mya) Patagonia and the Antarctic Peninsula were joined allowing various flora and fauna including land mammals to freely disperse between the continents. This route could have been used for migration of protoplatyrrhines from a variety of origins to South America (described in Houle, 1999).

Two scenarios consider an African origin of anthropoids. In the first, the protoplatyrrhines migrated from Africa through Asia to southeast Asia, Australia, and Antarctica eventually arriving in SA, while the other proposed a direct migration from Africa to Antarctica across the southwest Indian Ocean. An Asian anthropoid origin and Antarctic migration has also been proposed, assuming a migration of protocatarrhines to Africa and protoplatyrrhines through southeast Asia, Australia, and Antarctica arriving in SA. The paleoenvironment of Australia and Antarctica would have been suitable for the survival of a small land mammal, up until the late Eocene when continental Antarctica became glaciated and Australia experienced an abrupt drop in temperatures. This glaciation coincided with the opening of the Drake Passage, thus if protoplatyrrhines migrated to SA through Antarctica they must have arrived in SA prior to the late Eocene (37.2 mya) However due to the large aquatic barriers (2,600 km between Africa and Antarctica) and the number of water barriers encountered (2 between Asia and Antarctica) few support Antarctica as a viable migration route (Houle, 1999). Antarctic fossil deposits contain both marsupials and placental mammals however rodent or primate fossils have yet to be found there.

The most widely accepted migration route of primates to South America is a direct transatlantic route from Africa suggesting an African divergence of proto-platyrrhines and proto-catarrhines (Ciochon and Chiarelli, 1980b; Houle, 1999). In this scenario, early anthropoids diverged from Eurasian Omomyidae ancestors in the middle to late Eocene and inhabited Africa until a group, becoming proto-platyrrhines, journeyed to SA via island hopping and/or a floating island (i.e. rafting). Island hopping is unlikely, because both the Sierra Leone Rise and the Walvis Ridge were submerged

at that time, although, it is possible that the crest of the Rio Grande Rise may have remained subaerial (Houle, 1999). Migration via a raft or floating island is more probable, as it does not require the presence of intermediate landmasses. It has been hypothesized based on the predicted paleocurrents and paleowinds that a floating island could have crossed the middle Oligocene/ early Eocene Atlantic Ocean in approximately 7 to 14 days. This duration is predicted to have been a reasonable amount of time for the journey, during which the primates could have survived and maintained normal behaviors, such as feeding and reproduction, upon arrival (Houle, 1999). It is likely that the colonization of the Neotropics by platyrrhines and caviomorph rodents was synchronous, as there were only small windows in time during which the environment would have been conducive to a transoceanic journey. However given that caviomorphs split from their sister group phiomorphs almost 10 million years prior to the platyrrhine catarrhine divergence, we must then assume that caviomorph divergence occurred in Africa followed by a South American radiation (Poux et al., 2006).

Here we propose another possible route of migration for New World monkeys directly from Asia via a transpacific rafting event, to our knowledge this scenario has not been previously proposed. An Asian origin for anthropoid primates has been supported by several studies (Beard, 1998a; Bloch et al., 2007; Silcox, 2008; Springer et al., 2012; Stewart and Disotell, 1998). Recently, Springer and colleagues (2012) proposed that the mrca of anthropoid primates inhabited Asia. This finding raises the possibility in which New World monkeys dispersed to South America from Asia. We feel our biogeographic reconstruction most highly supports this route of migration to the Neotropics. Our data show that ancestral platyrrhines inhabited areas of western South

America (C and E), from 18 mya to at least 12.5 mya, when the ancestors of extant Cebinae migrated east (to region G). We also note that many early platyrrhine fossils are from localities in western SA (Hartwig and Meldrum, 2002). In addition, we show that the lineages of the three NWM families all arose in western South America (C and E), thus increasing the likelihood of this area as the origin of the platyrrhine radiation. The oldest known specimen from the primate fossil record in South America is *Branisella*, dated to 26 mya, from the La Venta fossil deposits in western Colombia (see Figure 4.1, (Fleagle and Tejedor, 2002)). Taken together, the data showing ancestral NWMs resided in western South America, the location of *Branisella*, and recent support for an Asian anthropoid origin, provide enough support to warrant further study into a direct migration of protoplatyrrhines from Asia to South America. Arguing against this scenario is the lack of early catarrhine or platyrrhine fossils in Asia, and the earliest catarrhines are known from the Fayum (Simons, 1997).

Although we can make inferences about the dispersals of extant NWMs (see below), we can only speculate on the likelihood of the above proposed scenarios describing migration of NWMs to South America. In order to pinpoint the region of South America that the protoplatyrrhines arrived, we would need to incorporate the geographic ranges of the early, now extinct, South American primate lineages. Using data from extant species we are limited to the past 18-25 million years (mrca of extant platyrrhines) in our biographic reconstruction. Including extinct lineages, in the form of fossils, would broaden our analysis and allow us to more accurately predict ancestral area ranges further into the past. Unfortunately, including our current knowledge of the platyrrhine fossil record would introduce several potentially problematic issues. First,

platyrrhine fossils have been identified based on morphological characteristics alone, and for many, their taxonomic placement is contested (reviewed in Fleagle and Tejedor, 2002; Harrison, 2002). Unknowingly placing the extinct taxa within the phylogenetic reconstruction would surely introduce inaccurate and misleading ancestral area results if the phylogenetic framework is incorrect. In addition, the Neotropics provide a harsh environment for fossil preservation, and the majority of mammal fossils in South America have been found only at a handful of rich reserves (e.g. La Venta, Colombia). These fossil rich sites provide us with few data points, from which to infer the ancestral ranges, of the extinct specimens found within them. Including this data would likely skew the results of an ancestral area reconstruction, as much more data is unknown, than is known and would be included. Due to the limited fossil record we are not yet able include extinct specimens in an analysis such as this; however, it is our hope, that as knowledge of extinct platyrrhines grows, we will soon be able to carry out such an analysis.

Neotropical primate radiations and diversification

Irrespective of direction of dispersal, once in South America, NWMs underwent multiple radiations and extinction events, resulting in the survival of a single ancestral lineage. From biogeographic reconstructions in the present study we suggest that the ancestor of extant platyrrhines likely lived in portions of Peru and Brazil bordered by the Amazon on the north, the Ucayali on the west, the Tapajos on the east and Madre de Dios and Guapore on the south (E). Around this time period, in the Early and Middle Miocene, western Amazonia existed as a large wetland with shallow lakes and swamps, known as the Pebas System (Latrubesse et al., 2010). The extensive rainforests that

previously existed became fragmented, and were displaced to the outer margins of the wetlands (Pons and De Franceschi, 2007). Along the Pacific coast the Andes continued their uplift, a process beginning 50 mya with the subduction, and breaking apart of the Pacific margin plate (Poulsen et al., 2010). This region of South America does not contain any known primate fossil deposits. *Branisella*, the oldest known platyrrhine fossil was identified south of this region in Salle, Bolivia and is dated to 26 mya (Fleagle and Tejedor, 2002; MacFadden, 1990). Based on our dating estimates, excluding *Branisella* from crown Platyrrhini, the mrca of platyrrhines existed, ~10 million years after the occurrence of *Branisella,* ~ 1,800 km north of Salle. There does not appear to be any geographic barriers to travel between Salle and the ancestral platyrrhine range during this time period. Several other primate fossils have been identified in South America during the Early Miocene, however none are near the predicted ancestral platyrrhine range. These fossils exist further south in Argentina (*Carlocbus*, *Donichocebus*, *Homunculus*, *Proteropithecia*, *Soriacebus*, and *Tremacebus*) and Chile (*Chilecebus*) (Fleagle and Tejedor, 2002). The wide geographic spread among Early Miocene New World primates could suggest multiple early radiations, resulting in only a single surviving lineage, that of extant platyrrhines. However, the 95% confidence interval for the mrca of extant platyrrhines is very broad when the node is not anchored using *Branisella* as a calibration point (13.04-40.16 mya). Thus it is possible that the Crown Platyrrhini diversified as early as the Late Eocene and could include all Neotropical primate fossils.

In the beginning of the Late Miocene rising sedimentation rates from the Andean basin to the Atlantic coast created the Acre system, a large body of water essentially

separating northern and southern Amazonia, and transitioning western Amazonia to a fluvial environment (Latrubesse et al., 2010). Just prior to this time period the mrca of extant Atelidae had dispersed north of this aquatic divide into what today are portions of Ecuador, Colombia, Peru and Brazil (i.e. OGU C). In addition to the mrca of extant Atelidae, this region was also inhabited by a number of extinct species. On the northern edge of this OGU is the rich fossil site of La Venta, Colombia - home to many Late Miocene fossil species, including several primates (*Stritonia*, *Cebupithecia*, *Nuciruptor*, *Neosaimiri*, *Laventiana*, *Aotus illigar*, *Mohanamico*, *Patasola*, *Lagonimico*, and *Micodon*) (Hartwig & Meldrum 2002). While a few of these fossils are thought to be related to atelids (*Stritonia*), others are assumed to be more closely related to pitheciids or the callitrichines, two groups who at this time were residing on the southern side of the Acre system. After the formation of the Acre system there is no evidence of movement of species across this barrier until < 5 mya when, due to a drop in global sea level and cooling, the Amazon River became fully established and heavily forested areas replaced the western Amazonian wetlands (Wesselingh and Salo, 2006).

South America's major geologic transitions were completed by the end of the Pliocene, with the formation of major Andean peaks and closing of the Panama isthmus, thus marking the start of the Great American Biotic Interchange (Hoorn et al., 2010). During this time primate fossils were identified in Brazil as well as in the Antilles. *Protopithecus brasiliensis* and *Caipora bambuiorum* were both identified in Brazil in the late Pleistocene. While the remains of *Paralouatta* was found in Cuba, *Xenothrix* in Jamaica, and *Antillothrix* in Hispaniola, all during the Quaternary (Macphee and Horovitz, 2002). In addition, this era was filled with species level migrations into these

areas, including: Central America and Mexico bounded on the south by the Magdalena (A), Peru bounded by the Ucayali on the east, the Maranon on the northwest and the Pacific coast (D), Brazil bordered by the Teles Pires on the east and the Tapajos on the west (H), and Argentina, Chile, Bolivia, Paraguay and a portion of Brazil bordered by the Madre de Dios and Guapore on the north, the Parana on the east and the coastline on the south and west (I). There is evidence that elements of the Middle America fauna were isolated through much of the Cenozoic (Werman, 2005). This isolation was due to climatic factors in the North and a water barrier in the South. The presence of platyrrhine fossils from the Quarternary of Hispaniola and Cuba is therefore intriguing as it suggests a pattern of stepping-stone dispersal that predates the closing of the Panamanian portal in the early Pliocene. Given the biogeographic scenarios we present, it is possible that platyrrhines began this dispersal to the Caribbean as early as 16 mya.

An aged based phylogenetic classification

Given the recent movement toward hyperinflation of species, and genus level taxonomy within New World monkeys (Rosenberger, 2012), we would like to take this opportunity to state our position in support of a phylogenetic, age based classification among all primates. In following the NWM taxonomy put forth by Groves (2005), and the age based classification scheme recommended by Wildman and Goodman (2004), we recognize 15 extant NWM genera (as shown in Table 4.1). Other classifications have supported the division of the marmosets into 4 genera, *Callithrix*, *Mico*, *Cebuella*, and *Callibella*, based primarily on distinct morphological variation. The division of the marmosets into multiple genera has been largely discussed in the literature from both a

morphological and molecular perspective (reviewed in Rylands et al., 2000; Schneider et al., 2012). Based on divergence times within the clade we place all marmosets within the genus *Callithrix*. Using this system we do not recognize *Mico*, *Callibella* or *Cebuella* as genera, as the crown age of *Callithrix* is only 4.13 mya (5.16 mya with the inclusion of *Branisella*), within the span (4-6 mya) that defines genera level taxonomy (Goodman et al., 1998; Wildman and Goodman, 2004).

A recent study looking at morphological and behavioral variation among capuchin monkeys suggested the division of the clade into two distinctive genera, *Cebus* (gracile capuchins) and *Sapajus* (robust capuchins) (Lynch Alfaro et al., 2012). However our dating analysis estimates the divergence date at the crown *Cebus* node at 3.74 mya (5.26 mya with the inclusion of *Branisella*), thus rejecting the classification of robust capuchin monkeys as an independent genus, *Sapajus*.

Our dating estimates infer the mrca of *Cacajao* and *Chiropotes* at 5.26 mya (excluding *Branisella* as a calibration point) and at 6.45 mya (including *Branisella*). These dates place the two genera on the cusp of being merged into a single genus, *Chiropotes*, as suggested by Wildman and Goodman (2004). Though we think that an age-based classification should be continued at the species level, we do not put forth a species level taxonomy, due to the limited species coverage in this analysis. Future divergence dating studies, with complete species and subspecies coverage, will be necessary to determine a classification scheme fitting with that proposed here at the genus level.

Conclusion

Using non-genic molecular markers we have shown that neutrally evolving DNA sequence is more phylogenetically informative than coding sequence, and in addition to increased dataset size the type of data is as important when attempting to resolve difficult phylogenies. Our resolved phylogenetic framework enabled us to estimate the divergence times of New World monkeys and reconstruct the ancestral geographic ranges in working toward revealing the tempo and mode of platyrrhine evolution.

Acknowledgments

The authors thank the anonymous reviewers for their comments and critique of the manuscript and Stephen D. Nash for providing beautiful Owl monkey illustrations. For the tissue samples we thank The Centro de Primatologia de Rio de Janeiro, The Brazil Centro Nacional de Primates, Dr. Caro-Beth Stewart, The University of Texas MD Anderson Cancer Center, The University of Pittsburg Medical Center, and The Center for Reproduction of Endangered species at the San Diego zoo. The authors would also like to acknowledge our deceased colleague Dr. Morris Goodman whose dedication to NWM phylogenetics continues to inspire our research. This research was funded by National Science Foundation Grants BCS-0751508 to DW and BCS-0751481 to SY. Natalie Jameson is funded by NSF Doctoral Dissertation Improvement Grant BCS-1061370.

CHAPTER 5

Genetic underpinnings of litter size reduction in Goeldi's monkey (*Callimico goeldii*)

Summary

Callitrichines differ from all other anthropoid primates, as they are the only clade to display obligate twinning. Within callitrichines, marmosets and tamarins produce 2-5 offspring per litter, however Goeldi's monkey, also a callitrichine, produces only a single offspring at parturition. The most parsimonious explanation for this difference in reproductive strategy is that twinning evolved on the callitrichine stem lineage and was subsequently lost on the Goeldi's monkey terminal lineage. Here we aim to shed light on the genetic underpinnings responsible for the change in reproductive strategy by first accessing the presence of chimerism in Goeldi's monkey, and secondly identifying genes undergoing adaptive evolution coincident with the phenotypic changes. Toward the detection of chimerism in Goeldi's monkey, both microsatellite genotyping and SRY amplification were unable to detect any sign of chimerism in our Goeldi's monkey study population (n=21). We identified 12 genes, involved in reproduction and ovulation, which have undergone adaptive evolution on lineages within the callitrichines. Genes known to alter ovulation rate, such as *BMPR1B* (bone morphogenetic protein receptor, type 1B), were identified as adaptively evolving on the callitrichine stem lineage. The known functions of the adaptively evolving genes as well as the absence of chimerism in

Goeldi's monkeys supports the presumption that Goeldi's monkeys have reverted to a state of mono-ovulation.

Introduction

Differences between organisms' life history traits, specifically those representative of reproductive strategy (i.e. gestation length, weaning age, and litter size) are part of what make species unique. Changes to these traits generally occur over long evolutionary timespans, e.g. between Orders (Promislow and Harvey, 1990). However, occasionally these changes occur within families or genera, thus resulting in a rapid shift in reproductive strategy. Rapid shifts in life history traits suggest an occurrence of dramatic environmental changes and/or an exceptionally valuable trade off for the species. Callitrichidae, a subfamily of New World monkeys, are a prime example of rapid shifts in reproductive strategy, as their reproductive strategies have be altered a minimum of two times in a ~10 million year period (Jameson at al., in review, See Chapter 4; Opazo et al., 2006; Perez et al., 2012).

Extant callitrichines include marmosets (*Callithrix*), tamarins (*Leontopithecus* and *Saguinus*), and Goeldi's monkeys (*Callimico*) (Groves, 2005). These monkeys have many synapomorphic traits unique only to their clade including; the presence of claws, reduced body size, alloparental care of offspring, short inter-birth intervals, two to four offspring per litter and absence of a 3rd molar (dental formula 2132/2132) (Ford, 1980; Ford and Davis, 1992; Hill, 1957; Martin, 1992; Wislocki, 1939). However, not all callitrichines share these unique features. *Callimico*, a single species genus (*Callimico goeldii*), stands out in that it is the only callitrichine genus to produce singleton offspring and have a 3rd molar (Altmann et al., 1988; Martin, 1992). Both of these traits, singleton

offspring and the presence of a 3rd molar, are ancestral to the callitrichine clade, while litter production and the presence of only two molars are derived traits shared by all but the single callitrichide genus (Table 5.1).

Table 5.1: Shared derived traits in Cebidae

	Claws	Mean body size (g)	No. of offspring	3rd molar	Mean interbirth interval (days)	Offspring care
Aotus	Absent	934	1	Present	381	Shared between mother and father
Saguinus	Present	469	2 - 4	Absent	234	Father and other females care for offspring
Leontopithecus	Present	600	2 - 4	Absent	311	Father and other females care for offspring
Callithrix	Present	309	2 - 4	Absent	183	Father and other females care for offspring
Callimico	Present	492	1	Present	270	Delayed care by father and other females
Saimiri	Absent	853	1	Present	900	Maternal care only
Cebus	Absent	2,804	1	Present	579	Maternal care only

Based on this morphological and physiological data many researchers have suggested that *Callimico* should group phylogenetically outside of the other callitrichines (Ford, 1986; Kay, 1990; Rosenberger, 1981, 1984). However, the presence of shared derived traits such as alloparental care, reduced body size, and short inter-birth intervals suggests otherwise. Modern molecular phylogenetic studies have conclusively determined that *Callimico* and *Callithrix* are sister genera forming a monophyletic clade sister to the tamarins (Canavez et al., 1999; Opazo et al., 2006; Pastorini et al., 1998; Perelman et al., 2011; Schneider et al., 2001; Wildman et al., 2009).

Here we focus on the rapid changes in reproductive strategy, specifically the variation in offspring number within the clade. Based on the phylogenetic structure of the clade, either three gains of multiple offspring litter production or a single gain and subsequent loss are required to achieve the current state (Figure 5.1). Given the similarity of the twinning processes in marmosets and tamarins and that it is more

parsimonious, we suggest a single gain of litter production on the callitrichine stem lineage and a single loss of twinning on the *Callimico* terminal lineage. The mechanism of multiple offspring production in marmosets and tamarins is distinctive in that it is initiated by multi-ovulation, the release of two to four ova per cycle (Davis, 2005; Tardif et al., 1993b; Wislocki, 1939). The mother is heavily influenced by environmental stressors when ovulating, varying the number of ova produced, and when carrying the fetuses. It is not uncommon for the mother to resorb one or more fetuses during the second half of gestation in response to the high nutritional and energetic demands of carrying multiple fetuses (Jaquish et al., 1996; Tardif et al., 2004; Windle et al., 1999). Beginning at day 19 of the 148 day gestation period chorionic fusion between the dizygotic twin fetuses occurs, allowing for the flow of blood, and therefore DNA, from twin to twin, resulting in chimeric offspring (Benirschke et al., 1962; Enders and Lopata, 1999; Gengozian et al., 1964; Moore et al., 1985; Wislocki, 1939).

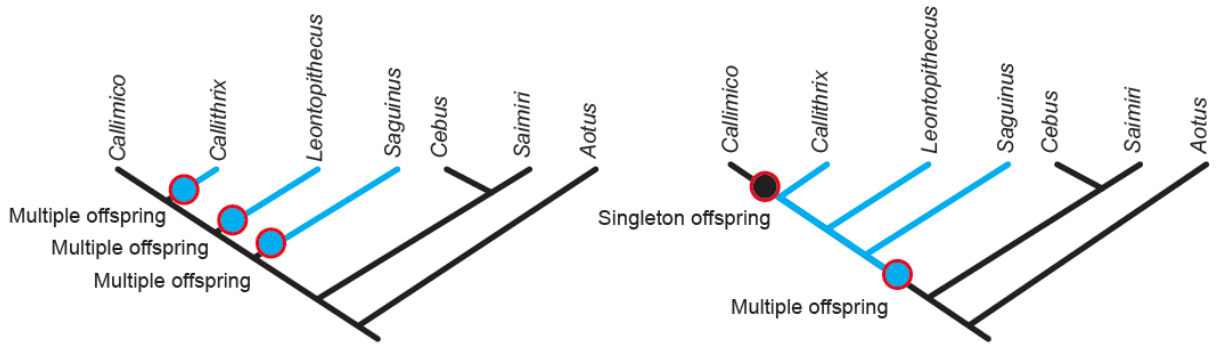


Figure 5.1: Evolution of offspring number variation
Alternate hypothesis showing the location of offspring number variation throughout the Cebidae family of New World monkeys. Blue branches represent lineages producing multiple offspring litters, and black branches represent lineages producing singleton offspring. Blue circle show a shift to an increase in number of offspring produced, and black circles show a shift to singleton offspring production.

Many of the callitrichine specific traits described above are direct effects of producing multiple offspring litters (e.g Benirschke, 1995). Morphological and behavioral

studies have observed the retention of traits such as diminutive body size and alloparental care on the *Callimico* lineage, suggesting that some aspects of the former reproductive strategy still remain. However, little is known about the retention of genetic traits associated with multiple offspring production. In this study, we aim to elucidate the genetic underpinnings of the phenotypic changes involved in litter size variation by investigating the timing and mechanism of litter size reduction and identifying adaptively evolving genes within callitrichines. We hypothesize that obligate fetus resorption could be the mechanism for litter size reduction in *Callimico goeldii*. Assuming fetus resorption occurred late in gestation, as with marmosets and tamarins, genetic information would have been exchanged between the fetuses, resulting in a chimeric singleton offspring. The presence of chimerism in *Callimico* offspring would imply that nutritional and energetic constraints forced the genus into the reduction of litter size. The absence of chimerism in *Callimico* offspring would suggest that the reduction in litter size occurs pre-ovulation, and is likely not linked with nutritional and energetic constraints. In addition, we hypothesize that genes involved in reproduction will show signs of adaptive evolution coincident with an increase or decrease of litter size. We investigate the specific roles these adaptively evolving genes play in reproduction to gain insight on the mechanisms involved in litter size variation and maintenance.

Here we present a genomic perspective on the phenotypic changes in reproductive strategy occurring within the callitrichines. We first aim to determine if *Callimico* singletons are chimeric. Using both microsatellite genotyping and SRY amplification we investigate the presence of non-self DNA in a colony of *Callimico goeldii*. Next we aim to investigate the history of genes involved in reproduction, and

specifically ovulation, during the evolution of the callitrichines. The identification of genes that are adaptively evolving coincident with changes in reproductive phenotype will provide further understanding of the physiological mechanisms involved in litter size alterations within primates.

Materials and Methods

Taxon sampling and collection of tissue samples

Included in our chimerism study is a population of 21 *Callimico goeldii* individuals and 6 *Callithrix* individuals of various species (Table 5.2). Frozen blood samples from the 21 *Callimico goeldii* individuals were obtained from a colony housed at the Brookfield zoo (Chicago Il.). These individuals are from three distinct family groups and vary in sex and age at tissue collection (Supplemental Figure S3). In addition, DNA from 6 *Callithrix* individuals was obtained from frozen blood samples acquired from The Brazil Centro Nacional de Primates and The Centro de Primatologia de Rio de Janeiro.

Tissue samples were also collected for RNAseq analysis (Table 5.2). Frozen tissue or fresh blood samples were obtained from *Aotus lemurinus* (liver), *Callimico goeldii* (blood), *Callicebus cupreus* (liver) and *Ateles fusciceps* (placenta).

Table 5.2: New World monkey tissue samples

Species	n =	Sex	Tissue	Source
Callimico goeldii	21	9 F, 12 M	Blood	The Brookfield zoo, Chicago Il
Callithrix jacchus	1	M	Blood	Goodman laboratory tissue sample collection
Callithrix flaviceps	2	1 F, 1 M	Blood	Centro de Primatologia de Rio de Janeiro
Callithrix Kuhlii	1	M	Blood	Centro de Primatologia de Rio de Janeiro
Callithrix geoffroyi	1	M	Blood	Centro de Primatologia de Rio de Janeiro
Callithrix penicillata	1	M	Blood	Brazil Centro Nacional de Primates
Aotus lemurinus	1	F	Liver	Goodman laboratory tissue sample collection
Callicebus cupreus	1	M	Liver	California National Primate Research Center
Ateles fusciceps	1	F	Placenta	Goodman laboratory tissue sample collection

DNA and RNA isolation

DNA was extracted from frozen blood samples using the Eppendorf gDNA Blood Mini kit (Hamburg Germany). Total RNA was extracted from the frozen tissue samples of *Aotus*, *Callicebus*, and *Ateles* using TRIzol Reagent (Invitrogen, Carlsbad, CA) and the Qiagen RNeasy Kit used in conjunction with the Qiagen RNase-Free DNase Set for clean up (Qiagen, Valencia, CA) according to the manufacturer's recommendations. Using the Leukolock Fractionation and Stabilization kit and protocol we extracted leukocyte RNA from a fresh blood sample from *Callimico goeldii*. The concentration and purity of the isolated DNA was tested using the NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE). The quality and quantity of all RNA samples was measured using the Agilent Bioanalyzer 2100.

Amplification and genotyping of microsatellite markers

In an attempt to detect non-self alleles in Goeldi's monkeys we genotyped 21 individuals with microsatellite loci known to identify non-self alleles (i.e. chimerism) in marmosets and tamarins (Escobar-Paramo, 2000; Nievergelt et al., 1998; Raveendran et al., 2008). The nine microsatellite loci were amplified in all 21 *Callimico* individuals as well as 6 *Callithrix* individuals. The microsatellites and accompanying primer sequences were obtained from previous publications (Escobar-Paramo, 2000; Nievergelt et al., 1998; Raveendran et al., 2008). PCR amplification was carried out using fluorescently labeled forward primers and unlabeled reverse primers. Products were visualized on agarose gels and purified using the Qiagen MinElute 96 PCR purification kit (Qiagen, Valencia CA.). Allele sizes were determined by separation of products on the ABI PRISM® 3130 Genetic Analyzer housed at the Research Technology Support Facility at

Michigan State University (East Lansing, MI). The fragment lengths were assigned using the Peak Scanner 1.0 (Applied Biosystems, Foster City, CA) program and a ROX GS500 size standard (ABI). Detection of >2 alleles in an individual, i.e. non-self alleles, indicated the presence of chimerism. If all alleles were represented in the individual's parental genotypes, than the individual was deemed to be chimeric.

Amplification of SRY region

We also attempted to detect chimerism by amplifying the Sex-determining Region Y (SRY) on the Y chromosome. PCR amplification of the appropriately sized amplicon in a female individual would suggest the presence a male twin's DNA, and thus chimerism. To test this we designed PCR primers from previously published Cebidae SRY sequence data (SRY_F1: 5'-GTAAGTTGCATTACAAAAGTTAAGG-3', SRY_R1: 5'-TCTTTGTAG TCAATGTTACCCG-3') (Moreira, 2002). We attempted to amplify this region in 9 female *Callimico*, as well as a male *Callimico*, a male *Callithrix*, and a female *Callithrix* as positive controls. The visualization of an appropriately sized amplicon on an agarose gel was deemed evidence of a present Y chromosome, and thus chimerism in a female.

Transcriptome sequencing

In order to obtain gene sequence data in a high throughput way, we produced transcriptomes for four New World monkey species without sequenced genomes. Total RNA (1.25 µg) from *Callimico goeldii*, *Aotus lemurinus*, *Callicebus cupreus*, and *Ateles fusciceps* was submitted to Expression Analysis (Durham, NC) per the company's instructions. RNA samples were prepared as paired end 2x100 bp libraries using the TruSeq RNA Sample Prep kit following the TruSeq Sample Preparation guide (Illumina,

San Diego CA). During this process unique adaptors were ligated to each sample allowing identification after multiplexing. All libraries were combined and each accounted for one third of a lane. RNAseq was performed using the Illumina HiSeq technology. By multiplexing the RNA samples we were able to obtain sufficient transcriptome coverage in a cost effective way.

Transcriptome assembly

Two methods were used for the assembly of the transcriptome data. The sequence data generated from RNAseq of *Callimico goeldii*, *Aotus lemurinus*, *Callicebus cupreus*, and *Ateles fusciceps* was assembled at Expression Analysis (Durham, NC) using Tophat2 and the marmoset genome as a guide for the assembly. In addition raw sequences were *de novo* assembled in house using the RNA-seq assembler Trinity (r2011-08-20).

Multiple sequence alignments

Here we investigate the evolutionary history of a list of genes (Table 5.3) that have been identified as being involved in reproduction. Although we are primarily interested in adaptive evolution occurrences within the callitrichine clade it is necessary to include additional species in the analysis. The inclusion of other primate and mammal species gives context to the callitrichines, by providing information on their ancestry. In addition, including more species breaks up long branches on the tree, thus preventing inaccurate outcomes that can occur when testing for selection. Thus we included 12 species with sequenced genomes. Transcripts of each of the genes of interest were obtained for each species from Ensembl (release 70).

Table 5.3: Genes involved in reproduction and ovulation

Gene symbol	Gene name	Function
ACVR1	Activine A receptor type I	Control of growth
ACVR2A	Activine A receptor type II	Control of growth
ACVR2B	Activine A receptor type IIB	Control of growth
AGT	Angiotensinogen	Vasculogenesis
ALMS1	Alstrom syndrome 1	Ovulation
BAMBI	BMP and activin membrane-bound inhibitor	Bone and cartilage development
BMP1	bone morphogenetic protein 1	Bone and cartilage development
BMP2	bone morphogenetic protein 2 preproprotein	Bone and cartilage development
BMP2K	BMP-2 inducible kinase isoform a	Bone and cartilage development
BMP4	Bone morphogenetic protein 4 preproprotein	Bone and cartilage development
BMP5	Bone morphogenetic protein 5 preproprotein	Bone and cartilage development
BMP6	Bone morphogenetic protein 6 preproprotein	Bone and cartilage development
BMP7	Bone morphogenetic protein 7 precursor	Bone and cartilage development
BMP8A	Bone morphogenetic protein 8A precursor	Bone and cartilage development
BMP8B	Bone morphogenetic protein 8B preproprotein	Bone and cartilage development
BMPR1A	Bone morphogenetic protein receptor, type IA	Bone and cartilage development
BMPR1B	Bone morphogenetic protein receptor type IB	Ovulation
BMPR2	Bone morphogenetic protein receptor type II	Bone and cartilage development
CCND2	Cyclin D2	Proliferation
CEBPB	CCAAT/enhancer binding protein (C/EBP), beta	Ovulation
CGA	Chorionic gonadotropin alpha chain	Reproductive physiology and endocrinology
CRP	C-reactive protein	Inflammation
CSH1	Chorionic somatomammotropin hormone 1	Control of growth
CSH2	Chorionic somatomammotropin hormone 2	Control of growth
CSHL1	Chorionic somatomammotropin hormone-like 1	Control of growth
CTSL1	Cathepsin L1	Ovulation
EGR1	Early growth response 1	Ovulation
ESR2	Estrogen receptor 2	Sexual development and reproductive function
FOXO3	Forkhead box O3	Apoptosis
FST	Follistatin	Reproductive physiology and endocrinology
FSTL1	Follistatin-like 1 precursor	Cell differentiaton
FSTL4	Follistatin-like 4 precursor	Cell differentiaton
GDF11	Growth differentiation factor 11	Cell differentiaton
GDF9	Growth differentiation factor 9	Cell differentiaton
GH1	Growth hormone 1- expressed in pituitary	Control of growth
GH2	Growth hormone 2- placentally expressed in human and macaque	Control of growth
GHR	Growth hormone receptor - binds GH	Control of growth
HAS1	Hyaluronan synthase 1	Ovulation
HAS2	Hyaluronan synthase 2	Ovulation
HIF1A	Hypoxia inducible factor	Hypoxia
HSD11B1	11-beta hydroxysteroid dehydrogenase	Cortisol regulation
IGF1	Insulin-like growth factor I	Fetoplacental growth
IGF1R	Insulin-like growth factor I receptor	Fetoplacental growth
IGF2BP1	Insulin-like growth factor 2 mRNA binding protein 1	Fetoplacental growth
IGF2BP2	Insulin-like growth factor 2 mRNA binding protein 2	Fetoplacental growth
IGF2BP3	Insulin-like growth factor 2 mRNA binding protein 3	Fetoplacental growth
IGF2R	Insulin-like growth factor II receptor	Fetoplacental growth
IGFBP2	Insulin-like growth factor binding protein 2	Fetoplacental growth
IGFBP4	Insulin-like growth factor binding protein 4	Fetoplacental growth
IGFBP7	Insulin-like growth factor binding protein 7	Fetoplacental growth

Table 5.3 Cont.

Gene symbol	Gene name	Function
IL1A	Interleukin 1, alpha	Inflamation
IL1B	Interleukin 1, beta	Inflamation
IL4R	Interleukin 4 receptor	Inflamation
INHBA	Inhibin, beta A	Reproductive physiology and endocrinology
ITIH2	Inter-alpha-trypsin inhibitor heavy chain 2	Ovulation
LEPR	Leptin receptor	Fetoplacental growth & trophoblast maturation
LGALS1	Galectin-1 - expressed in human fetal membranes	Immune response
LGALS3	Galectin 3 - expressed in human fetal membranes	Immune response
LGALS9	Galectin 9	Immune response
LHB	Luteinizing hormone beta subunit precursor	Reproductive physiology and endocrinology
LIF	Leukemia inhibitory factor	Immune response
MDM2	Mdm2 p53 E3 ubiquitin protein ligase homolog	Control of growth
MDM4	Mdm4 p53 binding protein homolog	Control of growth
MEST	Mesoderm specific transcript homolog (mouse)	Fetoplacental growth
MMP13	Matrix metallopeptidase 13	Cellular remodeling
MMP14	Matrix metallopeptidase 14	Cellular remodeling
MMP19	Matrix metallopeptidase 19	Cellular remodeling
MMP2	Matrix metallopeptidase 2	Cellular remodeling
MMP9	Matrix metallopeptidase 9	Cellular remodeling
MTHFR	Methylenetetrahydrofolate reductase	Cell processes
mTOR	Mechanistic target of rapamycin	Amino acid and glucose transfer
NRIP1	Nuclear receptor interacting protein 1	Hormone transcription regulation
PGR	Progesterone receptor	Ovulation
PIGF	Placental growth factor	Placental vasculogenesis
PLAT	Plasminogen activator, tissue	Cellular remodeling
PPARG	Peroxisome proliferator-activated receptor gamma	Adipocyte differentiation
PTGS2	Prostaglandin-Endoperoxide Synthase 2	Ovulation
SIRT1	Sirtuin 1	Gene regulation
SLC2A1	Glucose transporter 1	Amino acid and glucose transfer
SLC38A1	Solute carrier family 38	Amino acid and glucose transfer
SLC38A10	Solute carrier family 38	Amino acid and glucose transfer
SLC38A2	Solute carrier family 38	Amino acid and glucose transfer
SLC38A4	Solute carrier family 38	Amino acid and glucose transfer
SLC38A5	Solute carrier family 38	Amino acid and glucose transfer
SLC38A6	Solute carrier family 38	Amino acid and glucose transfer
SLC38A7	Solute carrier family 38	Amino acid and glucose transfer
SLC38A9	Solute carrier family 38	Amino acid and glucose transfer
TIMP1	TIMP metallopeptidase inhibitor 1	Cellular remodeling
TIMP2	TIMP metallopeptidase inhibitor 2	Cellular remodeling
TIMP3	TIMP metallopeptidase inhibitor 3	Cellular remodeling
TNF	Tumor necrosis factor α	Inflammation
TP53	Tumor protein p53	Apoptosis
USP7	Ubiquitin specific peptidase 7	Cell proliferation
VEGFA	Vascular endothelial growth factor A	Placental vasculogenesis
VEGFB	Vascular endothelial growth factorB	Placental vasculogenesis
WBP7	WW binding protein 7	Cell processes
WFIKKN1	WAP, Follistatin/Kazal, Immunoglobulin, Kunitz And Netrin Domain Containing 1	Cell differentiaton

The transcripts of our genes of interest were extracted from the 4 novel taxa (*Callimico goeldii*, *Aotus lemurinus*, *Callicebus cupreus*, and *Ateles fusciceps*) using two methods. Sequences of the genes of interest were identified in the Tophat transcripts by using the

transcript coordinates from the marmoset and human genomes. These gene sequences were also identified in the Trinity transcripts by BLAST of the human gene of interest against a database of the Trinity contigs (tblastx) and taking all hits that had 70% or greater sequence identity for 50 or more bases. Transcripts for each of the genes were aligned using MACSE v0.9b1 (Ranwez et al., 2011), a codon aware alignment software. For each species, each contig was scored against the human reference by calculating the ratio of amino acid matches (amino acid mismatches + amino acid matches), counting gaps in the contig as mismatches and counting insertions relative to the reference as neither a match nor a mismatch. The best scoring contig was selected as the representative for that species. Regions of poor alignment for each species were removed using a modified sliding window approach (Figure 5.2).

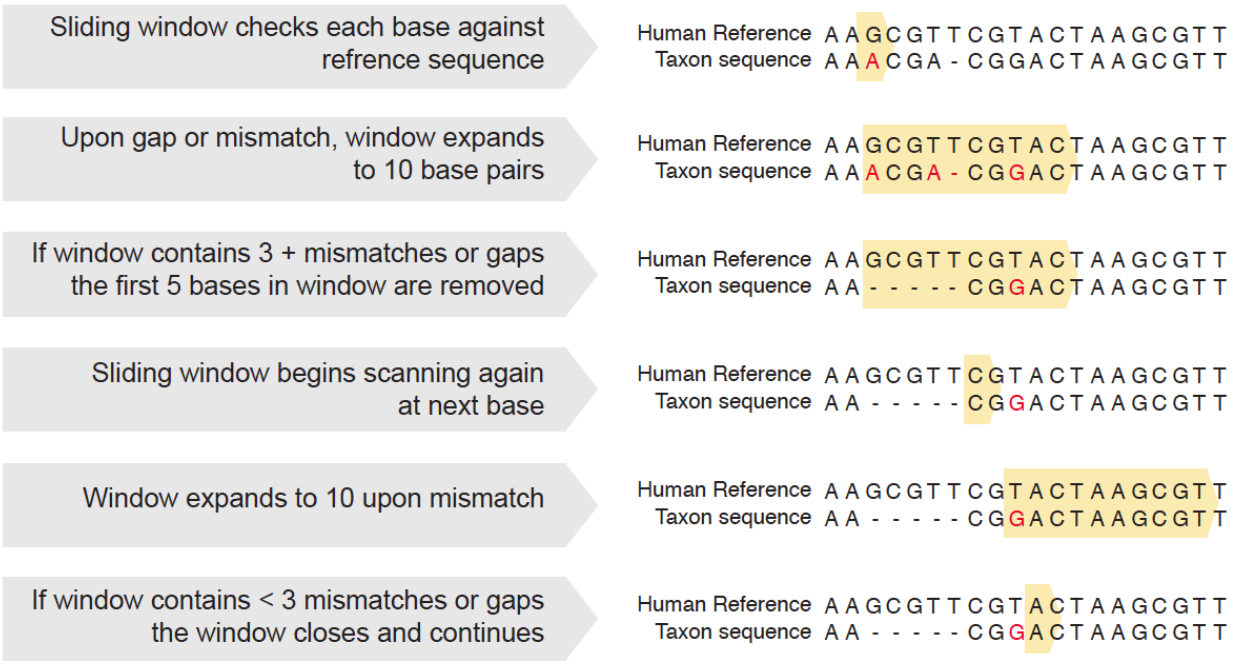


Figure 5.2: Sliding window sequence quality check
A sliding window of 10 base pairs was used to remove low quality sequence from our alignments.

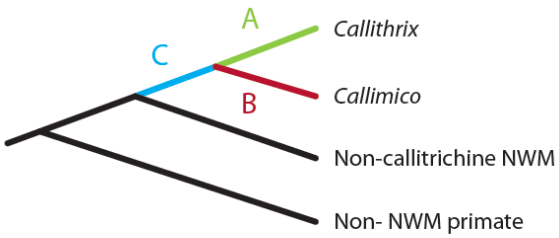
Starting at the first alignment position, each site was compared to the human reference. When a mismatch was detected, we checked a 10 base pair window starting at that site for identity to the reference. If more than three other mismatches were detected in the window (including gap sites as mismatches), a five base pair region at the beginning of the window was replaced with gap characters in the target sequence and scanning was resumed at the next amino acid site after the gap replacement.

Adaptive evolution of reproductive genes

The evolution of genes involved in reproduction was investigated using codeml in the PAML software package (Yang, 2007). Using this methodology we were able to determine if any of these genes are evolving adaptively (i.e. evolving significantly differently) on lineages where changes in litter size are thought to have occurred. This measure of gene evolution is determined as a ratio of the nonsynonymous substitutions per site over the synonymous substitutions per site, e.g. dN/dS. Both branch model tests, which allow dN/dS to vary among specified branches, and branch-sites model tests, which allow dN/dS to vary among codons along specified branches, were conducted.

Here, using the branch model we tested the terminal lineage of *Callimico* and *Callithrix* and the stem callitrichine lineage, allowing them to evolve at rates independent of the rest of the tree. Each gene was tested for evidence of adaptive evolution using four separate branch models including a null model, where all branches of the tree had the same dN/dS, and three alternate models, (1) where the callitrichine stem and terminal branches were allowed to vary independent of the rest of the tree, (2) where the terminal branches of *Callimico* and *Callithrix* were allowed to vary independently

from the stem, and from the rest of the tree, (3) where the terminal branches and stem were all allowed to vary independently from each other and from the rest of the tree (see Figure 5.3).



Models	Rates	Independent rate groups
Null	1ω	All branches
Alternate	2ω	**A**B**C** / Remaining branches
Alternate	3ω	**A**B / **C** / Remaining branches
Alternate	4ω	**A** / **B** / **C** / Remaining branches

Figure 5.3: Branch test models of adaptive evolution
The null and three alternate models of adaptive evolution, as analyzed in PAML under the branch test model. Each model allowed the rate of evolution to vary on branches A, B and C independently or in combination.

We compared all nested models to determine significance, using likelihood ratio tests (LRTs), taking into account the false discovery rate as determined by Q score to correct for multiple testing. The criteria for identifying a gene as adaptively evolving included (1) a significant preference ($p<0.05$) for an alternate model over the null following correction for multiple testing, and (2) a dN/dS ratio >0.05. When testing under the branch-sites model we analyzed the changes occurring at codons on specific branches of the tree. This model allowed for variation in rates on the *Callimico* terminal, *Callithrix* terminal, and the stem callitrichine lineages independently (Figure 5.4). All alternate models were compared with a null model to determine significance using likelihood ratio tests taking into account the false discovery rate as determined by Q score to correct for multiple testing. The criteria for identifying codon sites as adaptively evolving included (1) a significant preference ($p<0.05$) for an alternate model over the null following correction for multiple testing, and (2) a Bayes Empirical Bayes value of < 0.05 at the codon.

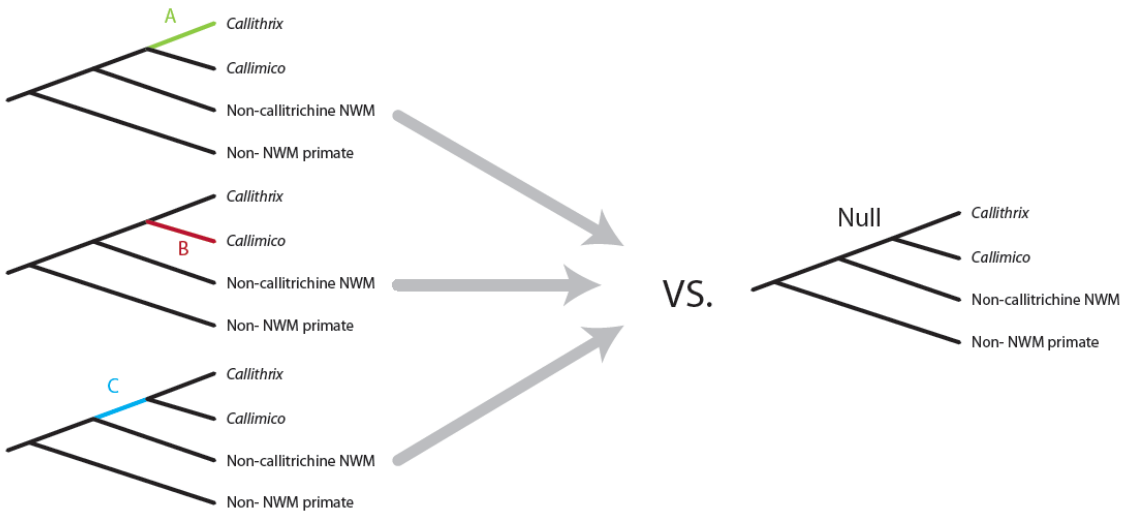


Figure 5.4: Branch-sites test model of adaptive evolution
The null and three alternate models of adaptive evolution, as analyses in PAML under the branch-sites model. Each alternate model allowed to rate of evolution at the codon sites to vary independently on branch A, B or C.

Results

Detection of chimerism in Callimico goeldii by microsatellite genotyping

In this study we hypothesize that late stage embryo resorption may be the mechanism of litter size reduction in Goeldi's monkey, thus resulting in the maintenance of a unique hallmark of marmoset and tamarins in Goeldi's monkey, chimerism. The presence of hematopoietic chimerism was investigated in 21 *Callimico* individuals and six *Callithrix* individuals using microsatellite genotyping (Table 5.4). Of the nine microsatellite loci amplified and genotyped none contained more than two alleles in any of the *Callimico* individuals, suggesting the absence of chimerism. As a positive control we genotyped the same nine microsatellites in six *Callithrix* individuals and found that all but one locus contained multiple (>2) alleles in at least one individual. Thus, confirming that hematopoietic chimerism is prevalent in *Callithrix* individuals at the loci tested.

Table 5.4: Microsatellite genotyping results

	Species	Studbook ID #	Mother	Father	In-house identifier	Allele A	Allele B	Allele C	Allele D
CAJA5, Raveendran 2008	*Callimico goeldii*	1729	NA	NA	Z1432	268	268		
	Callimico goeldii	974	NA	NA	Z1433	250	254		
	Callimico goeldii	2206	974	1729	Z1434	250	268		
	Callimico goeldii	2026	974	1729	Z1435	254	268		
	Callimico goeldii	1952	974	1729	Z1436	254	268		
	Callimico goeldii	2278	974	1729	Z1437	250	268		
	Callimico goeldii	1803	NA	NA	Z1438	248	258		
	Callimico goeldii	1810	NA	NA	Z1439	258	258		
	Callimico goeldii	2067	1810	1803	Z1583	258	258		
	Callimico goeldii	2201	1810	1803	z1584	248	258		
	Callimico goeldii	2368	1810	1803	z1585	258	258		
	Callimico goeldii	1805	NA	NA	z1586	258	258		
	Callimico goeldii	1806	NA	NA	z1587	258	258		
	Callimico goeldii	1991	1805	1806	z1588	258	258		
	Callimico goeldii	2066	1805	1806	z1589	258	258		
	Callimico goeldii	2124	1805	1806	z1590	258	258		
	Callimico goeldii	664	NA	NA	z1591	256	258		
	Callimico goeldii	1469	974	664	z1592	254	256		
	Callimico goeldii	1435	974	664	z1593	250	258		
	Callimico goeldii	1521	974	664	z1594	254	258		
	Callimico goeldii	1309	974	664	z1595	254	256		
	Callithrix jacchus	NA	NA	NA	z118	240	240		
	Callithrix flaviceps	NA	NA	NA	z1498	236	238		
	Callithrix penicillata	NA	NA	NA	z675	242	244		
	Callithrix flaviceps	NA	NA	NA	z703	240	242	236	238
	Callithrix kuhlii	NA	NA	NA	z704	242	244		
	Callithrix geoffroyi	NA	NA	NA	z705	242	244	246	248
CAJA10, Raveendran 2008	*Callimico goeldii*	1729	NA	NA	Z1432	186	197		
	Callimico goeldii	974	NA	NA	Z1433	183	185		
	Callimico goeldii	2206	974	1729	Z1434	185	197		
	Callimico goeldii	2026	974	1729	Z1435	183	197		
	Callimico goeldii	1952	974	1729	Z1436	183	186		
	Callimico goeldii	2278	974	1729	Z1437	183	197		
	Callimico goeldii	1803	NA	NA	Z1438	169	189		
	Callimico goeldii	1810	NA	NA	Z1439	169	189		
	Callimico goeldii	2067	1810	1803	Z1583	169	169		
	Callimico goeldii	2201	1810	1803	z1584	189	189		
	Callimico goeldii	2368	1810	1803	z1585	189	189		
	Callimico goeldii	1805	NA	NA	z1586	183	196		
	Callimico goeldii	1806	NA	NA	z1587	169	181		
	Callimico goeldii	1991	1805	1806	z1588	181	196		
	Callimico goeldii	2066	1805	1806	z1589	169	183		
	Callimico goeldii	2124	1805	1806	z1590	181	183		
	Callimico goeldii	664	NA	NA	z1591	181	181		
	Callimico goeldii	1469	974	664	z1592	181	183		
	Callimico goeldii	1435	974	664	z1593	181	185		
	Callimico goeldii	1521	974	664	z1594	181	185		
	Callimico goeldii	1309	974	664	z1595	181	185		
	Callithrix jacchus	NA	NA	NA	z118	178	182	184	
	Callithrix flaviceps	NA	NA	NA	z1498	182	184	196	198
	Callithrix penicillata	NA	NA	NA	z675	196	198	218	
	Callithrix flaviceps	NA	NA	NA	z703	190	194		
	Callithrix kuhlii	NA	NA	NA	z704	176	176		
	Callithrix geoffroyi	NA	NA	NA	z705	204	218		

Table 5.4 Cont.

	Species	Studbook ID #	Mother	Father	In-house identifier	Allele A	Allele B	Allele C	Allele D
CAJA11, Raveendran 2008	*Callimico goeldii*	1729	NA	NA	Z1432	249	255		
	Callimico goeldii	974	NA	NA	Z1433	255	255		
	Callimico goeldii	2206	974	1729	Z1434	255	255		
	Callimico goeldii	2026	974	1729	Z1435	255	255		
	Callimico goeldii	1952	974	1729	Z1436	255	255		
	Callimico goeldii	2278	974	1729	Z1437	255	255		
	Callimico goeldii	1803	NA	NA	Z1438	249	249		
	Callimico goeldii	1810	NA	NA	Z1439	255	257		
	Callimico goeldii	2067	1810	1803	Z1583	249	255		
	Callimico goeldii	2201	1810	1803	z1584	249	255		
	Callimico goeldii	2368	1810	1803	z1585	249	255		
	Callimico goeldii	1805	NA	NA	z1586	249	255		
	Callimico goeldii	1806	NA	NA	z1587	253	255		
	Callimico goeldii	1991	1805	1806	z1588	255	255		
	Callimico goeldii	2066	1805	1806	z1589	249	253		
	Callimico goeldii	2124	1805	1806	z1590	249	253		
	Callimico goeldii	664	NA	NA	z1591	241	257		
	Callimico goeldii	1469	974	664	z1592	241	255		
	Callimico goeldii	1435	974	664	z1593	241	255		
	Callimico goeldii	1521	974	664	z1594	255	255		
	Callimico goeldii	1309	974	664	z1595	241	255		
	Callithrix jacchus	NA	NA	NA	z118	235	253	255	
	Callithrix flaviceps	NA	NA	NA	z1498	233	239	239	
	Callithrix penicillata	NA	NA	NA	z675	247	251		
	Callithrix flaviceps	NA	NA	NA	z703	235	243	253	255
	Callithrix kuhlii	NA	NA	NA	z704	237	249	251	
	Callithrix geoffroyi	NA	NA	NA	z705	237	251	253	255
CAJA13, Raveendran 2008	*Callimico goeldii*	1729	NA	NA	Z1432	373	373		
	Callimico goeldii	974	NA	NA	Z1433	373	373		
	Callimico goeldii	2206	974	1729	Z1434	373	373		
	Callimico goeldii	2026	974	1729	Z1435	373	373		
	Callimico goeldii	1952	974	1729	Z1436	373	373		
	Callimico goeldii	2278	974	1729	Z1437	373	373		
	Callimico goeldii	1803	NA	NA	Z1438	371	371		
	Callimico goeldii	1810	NA	NA	Z1439	373	373		
	Callimico goeldii	2067	1810	1803	Z1583	371	373		
	Callimico goeldii	2201	1810	1803	z1584	371	373		
	Callimico goeldii	2368	1810	1803	z1585	371	373		
	Callimico goeldii	1805	NA	NA	z1586	371	373		
	Callimico goeldii	1806	NA	NA	z1587	373	373		
	Callimico goeldii	1991	1805	1806	z1588	373	373		
	Callimico goeldii	2066	1805	1806	z1589	371	373		
	Callimico goeldii	2124	1805	1806	z1590	371	373		
	Callimico goeldii	664	NA	NA	z1591	373	373		
	Callimico goeldii	1469	974	664	z1592	373	373		
	Callimico goeldii	1435	974	664	z1593	373	373		
	Callimico goeldii	1521	974	664	z1594	373	373		
	Callimico goeldii	1309	974	664	z1595	373	373		
	Callithrix jacchus	NA	NA	NA	z118	365	365		
	Callithrix flaviceps	NA	NA	NA	z1498	367	367		
	Callithrix penicillata	NA	NA	NA	z675	357	357		
	Callithrix flaviceps	NA	NA	NA	z703	355	365	367	
	Callithrix kuhlii	NA	NA	NA	z704	363	365		
	Callithrix geoffroyi	NA	NA	NA	z705	357	361		

Table 5.4 Cont.

	Species	Studbook ID #	Mother	Father	In-house identifier	Allele A	Allele B	Allele C	Allele D
CAJA17, Raveendran 2008	*Callimico goeldii*	1729	NA	NA	Z1432	433	433		
	Callimico goeldii	974	NA	NA	Z1433	425	437		
	Callimico goeldii	2206	974	1729	Z1434	433	437		
	Callimico goeldii	2026	974	1729	Z1435	425	433		
	Callimico goeldii	1952	974	1729	Z1436	425	433		
	Callimico goeldii	2278	974	1729	Z1437	433	437		
	Callimico goeldii	1803	NA	NA	Z1438	423	433		
	Callimico goeldii	1810	NA	NA	Z1439	423	433		
	Callimico goeldii	2067	1810	1803	Z1583	423	433		
	Callimico goeldii	2201	1810	1803	z1584	423	433		
	Callimico goeldii	2368	1810	1803	z1585	423	423		
	Callimico goeldii	1805	NA	NA	z1586	423	433		
	Callimico goeldii	1806	NA	NA	z1587	403	403		
	Callimico goeldii	1991	1805	1806	z1588	423	423		
	Callimico goeldii	2066	1805	1806	z1589	403	433		
	Callimico goeldii	2124	1805	1806	z1590	423	423		
	Callimico goeldii	664	NA	NA	z1591	427	433		
	Callimico goeldii	1469	974	664	z1592	425	427		
	Callimico goeldii	1435	974	664	z1593	425	427		
	Callimico goeldii	1521	974	664	z1594	427	437		
	Callimico goeldii	1309	974	664	z1595	425	433		
	Callithrix jacchus	NA	NA	NA	z118	347	367	370	
	Callithrix flaviceps	NA	NA	NA	z1498	361	381		
	Callithrix penicillata	NA	NA	NA	z675	369	373		
	Callithrix flaviceps	NA	NA	NA	z703	371	377	393	
	Callithrix kuhlii	NA	NA	NA	z704	353	387		
	Callithrix geoffroyi	NA	NA	NA	z705	387	397		
CAJA18, Raveendran 2008	*Callimico goeldii*	1729	NA	NA	Z1432	299	299		
	Callimico goeldii	974	NA	NA	Z1433	299	314		
	Callimico goeldii	2206	974	1729	Z1434	299	299		
	Callimico goeldii	2026	974	1729	Z1435	299	314		
	Callimico goeldii	1952	974	1729	Z1436	299	299		
	Callimico goeldii	2278	974	1729	Z1437	299	314		
	Callimico goeldii	1803	NA	NA	Z1438	305	314		
	Callimico goeldii	1810	NA	NA	Z1439	299	314		
	Callimico goeldii	2067	1810	1803	Z1583	305	314		
	Callimico goeldii	2201	1810	1803	z1584	314	314		
	Callimico goeldii	2368	1810	1803	z1585	305	314		
	Callimico goeldii	1805	NA	NA	z1586	314	314		
	Callimico goeldii	1806	NA	NA	z1587	308	314		
	Callimico goeldii	1991	1805	1806	z1588	314	314		
	Callimico goeldii	2066	1805	1806	z1589	314	314		
	Callimico goeldii	2124	1805	1806	z1590	308	314		
	Callimico goeldii	664	NA	NA	z1591	312	312		
	Callimico goeldii	1469	974	664	z1592	312	314		
	Callimico goeldii	1435	974	664	z1593	299	312		
	Callimico goeldii	1521	974	664	z1594	299	312		
	Callimico goeldii	1309	974	664	z1595	299	312		
	Callithrix jacchus	NA	NA	NA	z118	298	300		
	Callithrix flaviceps	NA	NA	NA	z1498	288	289	300	302
	Callithrix penicillata	NA	NA	NA	z675	294	296	296	
	Callithrix flaviceps	NA	NA	NA	z703	298	300	302	
	Callithrix kuhlii	NA	NA	NA	z704	289	295	302	
	Callithrix geoffroyi	NA	NA	NA	z705	298	300	302	

	Species	Studbook ID #	Mother	Father	In-house identifier	Allele A	Allele B	Allele C	Allele D
CJ12, Nievergelt 1998	*Callimico goeldii*	1729	NA	NA	Z1432	141	141		
	Callimico goeldii	974	NA	NA	Z1433	141	148		
	Callimico goeldii	2206	974	1729	Z1434	141	148		
	Callimico goeldii	2026	974	1729	Z1435	141	148		
	Callimico goeldii	1952	974	1729	Z1436	141	141		
	Callimico goeldii	2278	974	1729	Z1437	141	148		
	Callimico goeldii	1803	NA	NA	Z1438	148	148		
	Callimico goeldii	1810	NA	NA	Z1439	139	157		
	Callimico goeldii	2067	1810	1803	Z1583	139	148		
	Callimico goeldii	2201	1810	1803	z1584	148	157		
	Callimico goeldii	2368	1810	1803	z1585	148	157		
	Callimico goeldii	1805	NA	NA	z1586	139	141		
	Callimico goeldii	1806	NA	NA	z1587	139	141		
	Callimico goeldii	1991	1805	1806	z1588	141	141		
	Callimico goeldii	2066	1805	1806	z1589	139	141		
	Callimico goeldii	2124	1805	1806	z1590	139	141		
	Callimico goeldii	664	NA	NA	z1591	139	141		
	Callimico goeldii	1469	974	664	z1592	139	141		
	Callimico goeldii	1435	974	664	z1593	141	141		
	Callimico goeldii	1521	974	664	z1594	139	141		
	Callimico goeldii	1309	974	664	z1595	139	148		
	Callithrix jacchus	NA	NA	NA	z118	139	141	143	145
	Callithrix flaviceps	NA	NA	NA	z1498	153	156	158	
	Callithrix penicillata	NA	NA	NA	z675	139	141	141	
	Callithrix flaviceps	NA	NA	NA	z703	139	141	145	
	Callithrix kuhlii	NA	NA	NA	z704	139	143	153	
	Callithrix geoffroyi	NA	NA	NA	z705	141	141		
CJ13, Nievergelt 1998	*Callimico goeldii*	1729	NA	NA	Z1432	255	259		
	Callimico goeldii	974	NA	NA	Z1433	255	260		
	Callimico goeldii	2206	974	1729	Z1434	259	260		
	Callimico goeldii	2026	974	1729	Z1435	255	260		
	Callimico goeldii	1952	974	1729	Z1436	255	259		
	Callimico goeldii	2278	974	1729	Z1437	259	260		
	Callimico goeldii	1803	NA	NA	Z1438	251	259		
	Callimico goeldii	1810	NA	NA	Z1439	251	255		
	Callimico goeldii	2067	1810	1803	Z1583	251	259		
	Callimico goeldii	2201	1810	1803	z1584	255	259		
	Callimico goeldii	2368	1810	1803	z1585	251	251		
	Callimico goeldii	1805	NA	NA	z1586	255	257		
	Callimico goeldii	1806	NA	NA	z1587	251	257		
	Callimico goeldii	1991	1805	1806	z1588	257	257		
	Callimico goeldii	2066	1805	1806	z1589	255	257		
	Callimico goeldii	2124	1805	1806	z1590	257	257		
	Callimico goeldii	664	NA	NA	z1591	259	260		
	Callimico goeldii	1469	974	664	z1592	255	260		
	Callimico goeldii	1435	974	664	z1593	255	260		
	Callimico goeldii	1521	974	664	z1594	255	259		
	Callimico goeldii	1309	974	664	z1595	259	260		
	Callithrix jacchus	NA	NA	NA	z118	261	265	267	
	Callithrix flaviceps	NA	NA	NA	z1498	251	261	264	
	Callithrix penicillata	NA	NA	NA	z675	failed			
	Callithrix flaviceps	NA	NA	NA	z703	259	268		
	Callithrix kuhlii	NA	NA	NA	z704	245	255		
	Callithrix geoffroyi	NA	NA	NA	z705	264	269		

Table 5.4 Cont.

	Species	Studbook ID #	Mother	Father	In-house identifier	Allele A	Allele B	Allele C	Allele D
PEPC40, Escobar-Paramo 2000	*Callimico goeldii*	1729	NA	NA	Z1432	248	248		
	Callimico goeldii	974	NA	NA	Z1433	245	248		
	Callimico goeldii	2206	974	1729	Z1434	248	248		
	Callimico goeldii	2026	974	1729	Z1435	245	248		
	Callimico goeldii	1952	974	1729	Z1436	248	248		
	Callimico goeldii	2278	974	1729	Z1437	245	248		
	Callimico goeldii	1803	NA	NA	Z1438	248	252		
	Callimico goeldii	1810	NA	NA	Z1439	248	248		
	Callimico goeldii	2067	1810	1803	Z1583	248	252		
	Callimico goeldii	2201	1810	1803	z1584	248	252		
	Callimico goeldii	2368	1810	1803	z1585	248	252		
	Callimico goeldii	1805	NA	NA	z1586	248	250		
	Callimico goeldii	1806	NA	NA	z1587	245	254		
	Callimico goeldii	1991	1805	1806	z1588	250	254		
	Callimico goeldii	2066	1805	1806	z1589	250	254		
	Callimico goeldii	2124	1805	1806	z1590	245	250		
	Callimico goeldii	664	NA	NA	z1591	245	245		
	Callimico goeldii	1469	974	664	z1592	245	248		
	Callimico goeldii	1435	974	664	z1593	245	245		
	Callimico goeldii	1521	974	664	z1594	245	245		
	Callimico goeldii	1309	974	664	z1595	245	248		
	Callithrix jacchus	NA	NA	NA	z118	186			
	Callithrix flaviceps	NA	NA	NA	z1498	235			
	Callithrix penicillata	NA	NA	NA	z675	229	237		
	Callithrix flaviceps	NA	NA	NA	z703	failed			
	Callithrix kuhlii	NA	NA	NA	z704	229	240		
	Callithrix geoffroyi	NA	NA	NA	z705	failed			

Microsatellite result summary, ✓= Chimeric , ✗ = Not Chimeric

	CAJA5	CAJA10	CAJA11	CAJA13	CAJA17	CAJA18	CA12	CA13	PEPC40
Callimico goeldii	✗	✗	✗	✗	✗	✗	✗	✗	✗
Callithrix jacchus	✗	✓	✓	✗	✓	✗	✓	✓	✗
Callithrix flaviceps	✓	✓	✓	✗	✗	✓	✓	✓	✗
Callithrix penicillata	✗	✓	✗	✗	✗	✓	✓	✗	✗
Callithrix flaviceps	✗	✗	✓	✓	✓	✓	✓	✗	✗
Callithrix kuhlii	✗	✗	✓	✗	✗	✓	✓	✗	✗
Callithrix geoffroyi	✓	✗	✓	✗	✗	✓	✗	✗	✗

Detection of Chimerism in Callimico goeldii by SRY amplification

The SRY region (Sex-determining Region Y) of the Y chromosome is a highly distinctive sequence, which if amplified, indicates the presence of the Y chromosome. Amplification of the SRY region in female monkeys was done in an effort to detect male alleles and thus chimerism. SRY amplification was successful in a *Callimico goeldii*

male and a *Callithrix penicillata* male individual (Figure 5.5). Using the same procedure we were able to amplify the SRY region in a *Callithrix flaviceps* female individual, however attempts to amplify this region in 9 *Callimico goeldii* females were unsuccessful. These results show no male-female chimerism in our study group, and in combination with the genotyping results suggest the absence of any chimerism in our Goeldi's monkey study group.

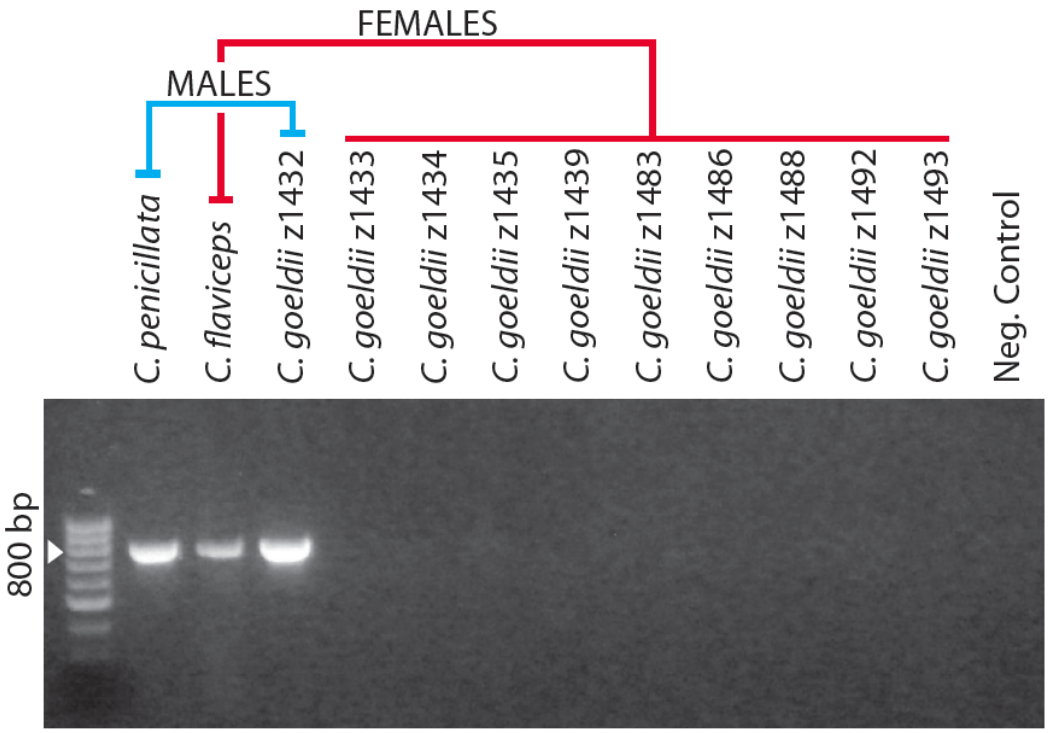


Figure 5.5: SRY amplification
Agarose gel electrophoresis of the Sex-determining Region Y (SRY) on the Y chromosome. The SRY was successfully amplified in a marmoset (*Callithrix penicillata*) and a *Callimico goeldii* male. The SRY was also successfully detected in a chimeric female marmoset (*Callithrix flaviceps*), however it was not detected in any of 9 *Callimico goeldii* females.

Assembled New World monkey transcriptomes

Due to the lack of available genomic data for these species we utilized moderate coverage RNAseq as a high throughput methodology for obtaining transcript data. We

sequenced total RNA from *Callimico goeldii* white blood cells, *Aotus lemurinus* liver tissue, *Callicebus cupreus* livertissue, and *Ateles fusciceps* placenta tissue generating a total of 20.6 Gb of data (Table 5.5). The transcriptomes were assembled *de novo* using the marmoset genome, the nearest relative to these species, as a guide (Tophat2) and also without a guide (Trinity). The combined data generated 121,997 contigs with a N50 of 2,523 base pairs (bp) from the *Callimico goeldii* data, 97,000 contigs with a N50 of 1,397 bp from the *Aotus lemurinus* data, 224,926 contigs with a N50 of 2,264 bp from the *Callicebus cupreus* data, and 232,772 contigs with a N50 of 1,903 bp from the *Ateles fusciceps* data. From this data we were then able to parse the transcriptome to obtain transcript sequence data for our genes of interest.

Table 5.5 RNAseq summary statistics

	Reads			Contigs			
Species	Total bases	No. mapped	Length	Total No.	Total length	Max length	N50
Aotus lemurinus	4,611,218,700	39,560,014	100	97,000	71,977,342	15,244	1,397
Ateles fusciceps	4,309,844,900	34,360,645	100	232,772	199,016,790	14,808	1,903
Callicebus cupreus	3,354,435,400	27,387,998	100	224,926	224,786,881	14,632	2,264
Callimico goeldii	8,360,180,100	76,606,413	100	121,997	132,445,144	24,538	2,523

Adaptively evolving genes in callitrichines

A total of 97 genes, known to be involved in reproduction and ovulation, were parsed from the transcriptomes and used to generate multiple sequence alignments of 16 mammalian taxa. Of the 97 genes of interest 76 multiple sequence alignments included all 16 taxa. Due to limitations in publicly available genomes and our transcriptomes the remaining 21 alignments were missing one or more taxa (Supplemental Table S9). The Codeml program within the PAML software package (Yang, 2007) was used to analyze the nucleotide multiple sequence alignments of each

gene. The use of branch model test and branch-sites model tests allowed us to look at the changes in the ratio of the nonsynonymous substitutions per site over the synonymous substitutions per site (dN/dS) on individual branches and groupings of branches, and along codon sites on the branches.

Our branch model analyses in PAML identified 12 genes that are evolving adaptively on lineages within the callitrichine clade (Table 5.6 and Figure 5.6). Under the alternate model (rate = 2ω), allowing the callitrichine stem and terminal branches to vary independently of the rest of the tree, we identified a single gene, *LEPR* (Leptin receptor) as having accelerated evolution. *LEPR* is evolving at a rate 2.36 fold faster within callitrichines than other taxa. The Leptin receptor is activated when bound by Leptin, an adipose-derived hormone shown to regulate aspects of food intake, energy balance, and fertility (Fruhbeck et al., 1998). When looking at the alternate branch model (rate = 3ω), allowing the callitrichine stem rate to vary independent of the terminal branches and of the rest of the tree, we identified 7 genes with accelerated evolution; *ALMS1* (Alstrom syndrome 1), *BMPR1B* (bone morphogenetic protein receptor, type 1B), *ESR2* (Estrogen receptor 2) *HAS1* (hyaluronan synthase 1), *GH1* (Growth hormone 1), *GH2* (Growth hormone 2), and *MMP9* (Matrix Metallopeptidase 9). Under this same model *BAMBI* (BMP and activin membrane-bound inhibitor homolog), a pseudoreceptor capable of limiting the range of TGF-beta family signaling in early embryogenesis (Onichtchouk et al., 1999), was identified as adaptively evolving on the terminal *Callimico* and *Callithrix* lineages. The alternate branch model (rate = 4ω), allowing the *Callimico* and *Callithrix* terminal lineages to vary independently of each other, the callitrichine stem lineage, and the rest of the tree, 4 additional adaptively

evolving genes were identified. On the *Callimico* terminal lineage *GH2*, a member of the somatotropin/prolactin hormone family, is again identified as adaptively evolving. On the terminal *Callithrix* lineage adaptive evolution was detected in *BMP8A* (Bone morphogenetic protein 8a) and *BMP8B* (Bone morphogenetic protein 8b) secreted signaling molecules that are part of the TGF-beta protein family, and *SLC38A4* (solute carrier family 38, member 4) a cationic and neutral amino acid transporter.

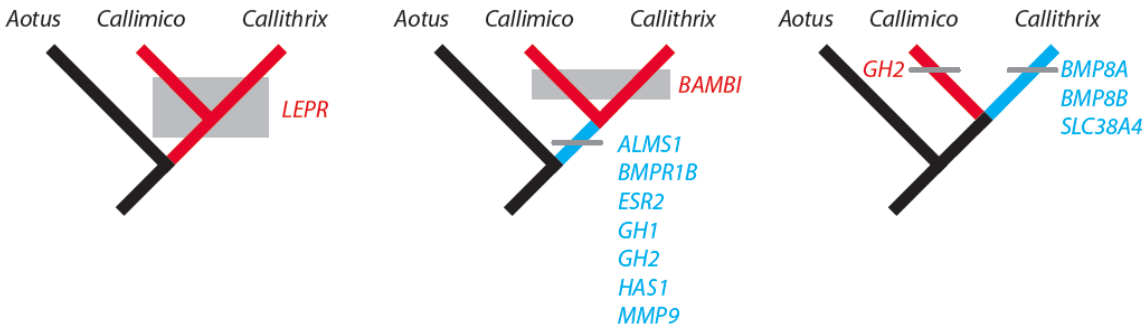


Figure 5.6: Genes undergoing adaptive evolution on callitrichine lineages
The 12 genes identified as undergoing adaptive evolution on callitrichine lineages through branch test analyses in PAML. A grey bar indicates which branch or group of branches each gene was identified as being adaptively evolving on.

Table 5.6: Genes undergoing adaptive evolution within callitrichines

		Null model	Alternate models					Branch A		Branch B		Branch C	
Adaptive evolution	Gene	1w	2w	3w	4w	LRT p value	Q score	N*dN	S*dS	N*dN	S*dS	N*dN	S*dS
Crown Callitrichine (A, B and C)	*LEPR*	-12005.2	-12002	-12002	-12002	p = 0.012	0.0792	0	0	7	5	23	16
Callitrichine stem (branch C)	*HAS1***	-5576.47	-5575.3	-5573.3	-5572.6	p = 0.026	0.0534					16	0
	ALMS1	-19827.6	-19823	-19823	-19823	p = 0.001	0.0673					36	15
	BMPR1B	-5604.27	-5600.6	-5600.3	-5600.3	p = 0.009	0.0895					1	0
	ESR2	-5411.83	-5407.9	-5403.8	-5403.8	p = 0.002	0.0673					16	0
	MMP9	-7668.5	-7668.5	-7665	-7664.9	p = 0.004	0.0896					4	0
	*GH1**	-1984.66	-1979.3	-1979.3	-1979.3	p = 0.005	0.0410					39	20
	*GH2**	-1001.38	-1001.4	-1000.9	-996.13	p = 0.001	0.0135					6	2
Callitrichine terminal ineages (A and	*BAMBI*	-2436.81	-2436.5	-2432	-2432	p = 0.001	0.0673	1	0	2	0		
Callithrix terminal (branch A)	*BMP8A*	-5688.83	-5675.6	-5688.8	-5667.6	p = 0.000	0.0066	35	13				
	BMP8B	-5311.74	-5311.5	-5311.5	-5289.2	p = 0.000	0.0000	31	14				
	SLC38A4	-6170.38	-6163	-6163	-6162.6	p = 0.000	0.0103	1	0				
Callimico terminal (branch B)	*GH2*	-1001.38	-1001.4	-1000.9	-996.13	p = 0.001	0.0135			6	0		

Branch A- *Calllimico* terminal; Branch B- *Callithrix* terminal; Branch C- Callitrichine stem
*Callitrichine stem to *Ateles*, **Callitrichine stem to *Callicebus*

The branch-sites model analyses were unable to identify any codon sites evolving adaptively on lineages within the callitrichine clade (Table 5.7 and Figure 5.7). However, three genes did produce significant LRT scores and contained sites undergoing accelerated evolution. On the callitrichine stem lineage (terminating at *Callicebus*) *HAS1*, Hyaluronan Synthase, had a significant LRT score after correcting for multiple testing. Within the gene five sites were undergoing accelerated evolution, one in the second and third codon each, and a cluster of three in the fourth codon (BEB p value = 0.499-0.129). On the *Callimico* terminal branch LHB (Luteinizing Hormone Beta Polypeptide) had a significant LRT score, however the statistical significance was lost after correcting for multiple testing. A single codon, site 107, was detected as having undergone accelerated evolution (BEB p value = 0.337). Matrix metallopeptidase 9 (MMP9) had a significant LRT score on both the *Callimico* and the *Callithrix* terminal lineages, however the statistical significance was lost after correcting for multiple testing. On the *Callithrix* terminal lineage three codons were found to be undergoing accelerated evolution (BEB p value = 0.499-0.129), and on the Callimico lineage a single codon was identified as undergoing adaptive evolution (BEB p value = 0.046).

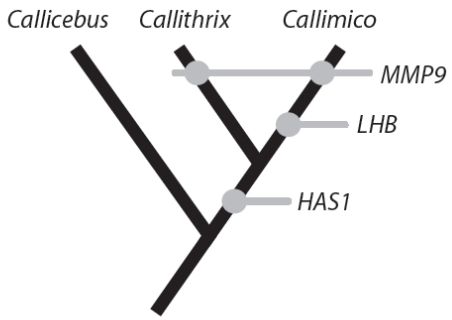


Figure 5.7: Genes with sites undergoing adaptive evolution on callitrichine lineages
Three genes were identified as containing sties undergoing accelerated evolution on callitrichine lineages through branch-sites tests in PAML. A grey bar indicates which branches each gene had sites that were identified as being under accelerated evolution.

Table 5.7: Codon sites undergoing accelerated evolution

Gene	Branch	Alternate	Null	p value	Site	AA	p value
HAS1	callitrichine stem	-5530.033	-5535.966	**< 0.000***	297	Gly	0.499
					324	Pro	0.129
					397	Arg	0.131
					403	Leu	0.138
					406	Asp	0.168
MMP9	Callithrix terminal	-7547.367	-7550.485	0.007	158	Leu	0.424
					441	GLn	0.337
					580	Lys	0.260
MMP9	Callimico terminal	-7547.741	-7549.821	0.024	241	Asn	**0.037****
LHB	Callimico terminal	-1531.383	-1534.622	0.006	107	Ala	0.337

* indicates p value that, following testing for multiple hypothesis, was significant
** indicated p value significant at 0.05

Discussion

Mechanisms underlying reversion to singleton offspring

Callitrichines, excluding *Callimico*, display multi-ovulation, the ability to produce 2-5 ova per reproductive cycle (Hill, 1932; Wislocki, 1939). The number of eggs produced is highly dependent on the mothers nutrition and stress levels. Marmosets in captivity (i.e. with low stress and sufficient nutrition) are more likely to produce 3-5 eggs per cycle, while those in the wild produce less (2-4 eggs) and will often reduce offspring number mid-gestation in response to changing environmental factors (Jaquish et al., 1996; Kirkwood, 1983; Tardif and Jaquish, 1997). Given that Goeldi's monkeys produce singleton offspring regardless of environmental factors, it has been assumed that the change in reproductive strategy was mediated via a reversion to the ancestral state of mono-ovulation (Tardif and Ross, 2009). However, to our knowledge, studies have yet to be conducted confirming the assumption that only a single ova is produced.

The current conservation status of *Callimico goeldii* is listed as 'vulnerable' and the population has been declining over the past decades, making the species difficult to

study both in the wild as well as in captivity (Comejo, 2008). In this study we took a non-invasive approach to infer (1) if *Callimico goeldii* exhibits mono-ovulation and if so (2) is embryo resorption the mechanism of litter size reduction. Our data show a complete absence of chimerism in our study population, thus suggesting late stage embryo resorption is not likely, also further supporting the assumption of mono-ovulation within the genus. We hypothesized that a prolonged environmental stressor, such as a reduction in feeding resources, necessitated conservation of maternal energy, thus the shift to obligate embryo resorption in *Callimico*. Although we see no evidence of embryo resorption in *Callimico*, it remains likely that environmental factors affecting the mother's nutritional status would have played a role in decreasing the litter size. As discussed above, callitrichines are unique in their ability to vary the number of ova produced based on maternal nutritional status. Therefore we suggest that mono-ovulation occurs in *Callimico goeldii* and is the result of a past response to a prolonged environmental stressor.

Footprints of a former reproductive strategy in Goeldi's monkey

The transformation of an organism from producing singleton to multiple offspring litters carries with it many more changes than simply the number of offspring produced at parturition. In the twinning callitrichines, many physiological, immunological, and behavioral traits can be seen as direct adaptations to their altered reproductive strategy (Benirschke et al., 1962). Although Goeldi's monkey does not produce multiple offspring litters, the genus has evolved from ancestors that were capable of twinning, and thus many of these adaptations have remained. Like all other callitrichines, Goeldi's monkey has retained a diminutive size, increased alloparental care for offspring, and a short

inter-birth interval. The retention of these traits suggests they have continued to provide an evolutionary advantage regardless of litter size.

Callitrichines contain the smallest of the New World monkeys, with the pygmy marmoset weighing only an average of 128g (Ford and Davis, 1992). The diminutive size of these primates is a derived trait, a secondary reduction in body size (Ford and Davis, 1992). Many hypotheses have suggested that twinning evolved in response to the decreased size of the callitrichines, as a way to compensate for their increased vulnerability to predators, or due to restrictions of passing a large brained infant through a small pelvic inlet (Eisenberg, 1981; Leutenegger, 1979; Tardif and Ross, 2009). However, since the reversion to singleton offspring, Goeldi's monkeys have maintained their decreased body size and are capable of producing enough offspring to maintain their population despite predation and successfully birthing their offspring. Regardless of the cause of the secondary reduction in body size, the fact that the adaptation has remained in all callitrichine species suggests it carries with it additional benefits. Callitrichines exist in habitats that go unutilized by other primates, preferring lower forest strata, secondary growth and vertical forests (Ferrari, 1993; Ferrari and Lopes Ferrari, 1989). The ability to live in these ranges is facilitated by their small size and preference for gum feeding. Had *Callimico goeldii* become a larger sized primate, it would have been necessary for the species to explore alternate environmental niches and food resources (Eisenberg, 1981; Ford and Davis, 1992).

Goeldi's monkey, along with marmosets and tamarins do not undergo lactational anovulation as most primates do, and remain fertile while lactating (Goldized, 1987). This often results in pregnancy within weeks of parturition, and therefore the short inter-

birth intervals observed in callitrichines. For the mother, simultaneous gestation and lactation is a major drain on energetic and nutritional resources. However, due to the availability of high calorie foods year round, such as gums, these elevated energy demands can be met (Nash, 1986; Sussman and Kinzey, 1984). Callitrichines exist in single cooperative breeding groups, made up of a monogamous male and female mating pair, their offspring, and subordinate non-breeding adults (Goldized, 1987). Because there is only a single breeding female, the short inter-birth intervals as well as multiple offspring litters allow for greater population growth. Although Goeldi's monkeys produce singletons, it is likely that they are able to maintain population growth because of these short inter-birth intervals.

A significant amount of time and energy is required by the primate mother for transporting her infant in the first months of life (Tardif et al., 1993a). Callitrichine pregnancies during lactation are greatly facilitated by the large amount of paternal and alloparental care provided to the infant. Within two days after birth, wild marmosets and tamarins have been observed transferring the infants to the father for transportation, while Goeldi's monkey mothers wait until week two before transferring the infant to the father (reviewed in Ross et al., 2010). In addition to paternal care, siblings and non-breeding females in the group also care for the young offspring, displaying alloparental care. The willingness of callitrichine helpers to participate in allocare can, similarly with most animals, be attributed to inclusive fitness and kin selection. The presence of chimerism increases the relatedness of siblings to each other as well as aunts and uncles to nieces and nephews. Therefore, following the formula for altruism set forth by Hamilton (rb>c) (Hamilton, 1964), the increase in relatedness in chimeric callitrichines

skews the equation such that the relatedness of the individuals (r) multiplied by the reproductive benefit of the recipient (b) is most often greater than the reproductive cost to the helper (c). It is likely that the increased relatedness among chimeric families is heavily responsible for the amount of allocare displayed in callitrichine populations (Haig, 1999). However, as determined by this study, *Callimico goeldii* does not show any signs of chimerism, yet cooperative breeding and allocare have remained part of the social system. Although *Callimico goeldii* mothers do delay in allowing paternal care, it's likely a result of singleton offspring, as immediate help in transporting the offspring is not necessary. Additionally, we point out that paternal care of offspring is also present in the genus *Aotus* (see Table 5.1). The phylogenetic position of *Aotus*, as sister to the callitrichine clade suggests that paternal care began prior to multiple offspring litters. In this case paternal, and possible alloparental, care was not a consequence of producing multiple offspring, but a factor that facilitated the production of litters.

Role of Leptin in litter size

Leptin is an adipose-derived circulating hormone shown to regulate aspects of food intake, energy balance, and fertility. Leptin's interactions occur through binding the leptin receptor, a class I cytokine receptor with widespread distribution throughout tissues, including the ovaries (reviewed in Fruhbeck et al., 1998). The analyses presented in this study found the leptin receptor gene (*LEPR*) to be adaptively evolving on all callitrichine lineages tested. *LEPR* has 23 nonsynonymous substitutions (N*dN) and 16 synonymous substitutions (S*dS) on the stem callitrichine branch, and 7 N*dN and 5 S*dS on the *Callimico* terminal branch. Interestingly, the *Callithrix* terminal lineage does not contain any substitution (N*dN=0, S*dS=0). Accelerated evolution of *LEPR*

appears on the lineages where reproductive strategy was altered (the gain of twinning, and loss of twinning), and not on the lineage where the reproductive strategy was maintained suggesting that this evolution provided the (partial) genetic substrate for change in this phenotypic trait.

Leptin was first described as the obesity gene, because mice that were resistant or deficient in leptin became obese (Campfield et al., 1995; Coleman, 1973; Halaas et al., 1995; Pelleymounter et al., 1995; Zhang et al., 1994). In addition to being obese, the leptin deficient mice (*ob/ob*) were infertile, however the condition was reversible with the treatment of leptin (Chehab et al., 1996; Hamann and Matthaei, 1996; Mounzih et al., 1997). In mice increasing the circulating leptin can trigger early puberty and reproductive tract maturation, and in humans a rise in leptin concentration is observed prior to puberty (Chehab et al., 1997; Garcia-Mayor et al., 1997; Mantzoros et al., 1997). Leptin, acting through its receptor, triggers the release of gonadotropin releasing hormone 1 from hypothalamic cells causing the release of luteinizing hormone (*LH*) and follicle stimulating hormone (*FSH*) (Yu et al., 1997). The follicle stimulating hormone receptor has been implicated in control of timing of human birth (Plunkett et al., 2011). Studies on multiple ovulation in sheep have shown that the concentration of *FSH* during the time of follicular selection controls the quantity of eggs released (reviewed in Montgomery et al., 2001). Alteration to the structure of the leptin receptor would lead to changes, either an increase or decrease, in the binding affinity for leptin and therefore changes in the concentration of *FSH* present during follicular selection. Given the role of leptin in ovulation and the presence of adaptive evolution in *LEPR* coincident with

changes in reproductive strategy, it is likely that *LEPR* may contribute to the mechanism of litter size variation in callitrichines.

The TGF-β superfamily and ovulation

The genes that make up the transforming growth factor beta (TGF-β) superfamily are cell regulatory proteins that are all structurally related. Contained within this family are the bone morphogenetic proteins (BMPs), a group of 20 growth factors. Our analysis found adaptive evolution in three genes within this family; *BMP8A*, *BMP8B*, and *BMPR1B*. Studies conducted looking at the expression of *BMP8A* and *BMP8B* found expression of the two highly similar genes in the decidua cells and placenta of mice respectively (Zhao and Hogan, 1996). The function of *BMP8A* in the mouse decidua cells is thought to promote the survival, proliferation and/or differentiation of the uterine cells, while the function of *BMP8B* in the placenta is proposed to induce proliferation of labyrinthine trophoblast cells (Shimasaki et al., 2004; Zhao and Hogan, 1996). Given the timing and location of the genes expression in mice, *BMP8A* and *BMP8B* appear to function as mediators to the maternal fetal interface. Our analyses showed that both *BMP8A* and *BMP8B* have undergone accelerated evolution on the *Callithrix* terminal lineage. It seems likely that these genes could be influencing the maternal fetal interface, or could play a role in facilitating the chorionic fusion among fetuses. The *BMPR1B* gene, or *FecB* locus, has been widely studied for its effects on the ovulation rate in sheep, mutations at the *FecB* locus ($FecB^B$) cause an increase in ovulation rate (see Bindon, 1984). Sheep containing the $FecB^B$ allele do not undergo the last stage of granulosa cell doubling and instead progress directly into a differentiation cascade (McNatty and Henderson, 1987; Montgomery et al., 1992).

Therefore, these sheep produce an increased number of smaller follicles. The *FecB*^{B} mutation was found not to act by increasing gonadotrophic stimulation on the follicles, but by increasing the sensitivity of the follicular cells to gonadotrophic stimulation (Campbell et al., 2003). Our data detected adaptive evolution of *BMPR1B*, the gene containing the *FecB* locus, on the callitrichine stem lineage. The changes in the evolution and therefore function of this gene, coincident with the gain of multiple offspring litter production in callitrichines, points toward a key direction for further studies of the mechanism of ovulation rate increase in callitrichines.

Conclusions

The molecular mechanisms underlying changes in the reproductive strategies of callitrichines have yet to be fully determined. The current study provides insight into the molecular mechanisms and establishes a basis from which to continue these studies. We determined that (1) chimerism is not present in *Callimico goeldii*, supporting the assumption of mono-ovulation in the species and (2) genes involved in ovulation rate are adaptively evolving coincident with changes in litter size. Although our data cannot confirm mono-ovulation in *Callimico goeldii*, ultrasonography techniques currently used on marmosets could be utilized to verify our finding. Future studies on the time and specific location of gene expression as well as functional studies of the genes found to be undergoing adaptive evolution will provide much more insight into the roles of theses genes predicted.

Acknowledgments

This research was funded by National Science Foundation Grants BCS-0751508 to DW. Natalie Jameson is funded by NSF Doctoral Dissertation Improvement Grant

BCS-1061370. We thank The Brookfield zoo for the kind gift of multiple tissues samples. Tissue samples for this research were also obtained from Centro de Primatologia de Rio de Janeiro, The Brazil Centro Nacional de Primates, Dr. Caro-Beth Stewart, The University of Texas MD Anderson Cancer Center, The University of Pittsburgh Medical Center, California National Primate Research Center (NIH-RR00169), and The Center for Reproduction of Endangered species at the San Diego zoo for tissue samples.

CHAPTER 6

Conclusions

Reproduction is a fundamental requirement for organic evolution, thus modifications to reproductive strategy may significantly alter the organisms evolutionary course. The aim of my research was to integrate phenotypic and genotypic character evolution in order to determine the molecular underpinnings of the varied reproductive strategies in NWMs. Toward this end, I first set out to develop a highly supported phylogenetic framework. Using this foundation I was then able to investigate the biogeographic history of NWMs by linking the predicted ancestral ranges and divergence times to key paleogeographic and paleoclimactic events. Focusing in on the callitrichide clade of NWMs, I used a comparative genomics approach to identify adaptively evolving genes that could have played a role in the alteration of reproductive strategy. Lastly, through the detection of twin-twin chimerism, I sought to determine the timing (pre or post-fertilization) of litter size reduction in the Goeldi's monkey.

In Chapter 2, I set out to determine the sister taxa to anthropoid primates. Through large scale phylogenomic analysis I tested the assumption that Prosimians (Tarsiers and strepsirrhines) form a monophyletic clade to the exclusion of anthropoid primates. The phylogenetic positioning of these three clades has been widely studied (Hasegawa et al., 1990; Murphy et al., 2001; Chatterjee et al., 2009, Pocock, 1918; Baba et al., 1982; Goodman et al., 1998; Zietkiewicz, et al., 1999; Schmitz et al., 2001, Andrews et al., 1998; McNiff and Allard, 1998). However, little consensus has been

reached. Taking advantage of the newly released *Tarsius syrichta* genome, our analysis determined the genus *Tarsius* to be the sister taxa of anthropoid primates. Shimodaira-Hasegawa topology testing not only supported this arrangement as the most likely, but also found it to be statistically more likely than the other possible topologies. With the knowledge that *Tarsius* is the sister taxa to anthropoid primates, we are now able to correctly infer the ancestral anthropoid state of any primate trait. Interestingly, we note that *Tarsius*, a genus often categorized as “primitive”, has undergone a greater amount of protein-coding substitutions than any other anthropoid taxa sampled in the study. Given the amount of genomic evolution on this branch, we reject the nomenclature of “higher” and “lower” primates.

Although whole genome sequencing has become more prevalent, there are still many primate species with very limited genomic sequence data available. NWMs fall into this category. In Chapter 3 we worked to overcome this lack of data by generating whole genome shotgun libraries from spider monkey, owl monkey and Uakari, subsequently providing us with 3,154 individual sequences containing nearly 2.3 million basepairs. Annotation of these sequences to the human genome determined that 57% of the sequences were intergenic, 38% were in introns, 2.2% were within 2 kilobases of a transcription start site, and 0.2% were within exons. We were also able to identify 664 sequences with marmoset orthologs and 659 sequences with human orthologs. The genomic data generated here exist in a variety of genic and non-genic regions and can be used to develop molecular resources such as phylogenetic or microsatellite markers. In the ensuing section we have demonstrated the utility of this genomic resource.

In Chapter 4, we sought to infer the phylogenetic history of NWMs and determine how it may have been influenced by environmental factors. A phylogenetic framework was built using a data matrix containing 36 NWM taxa and 40,986 basepairs of non-genic sequence date. Analysis of this data matrix produced a highly supported phylogeny, with 88% of the nodes fully supported. We report that non-genic markers do in fact produce a strong phylogenetic signal (Peng et al., 2009; Wildman et al., 2009), and when used in large quantities, can be more informative than traditional phylogenetic markers. Using the known ranges of extant NWMs we performed ancestral area reconstructions to infer ancestral ranges and the clades historical biogeography. This data gave us insight into the geographic paths taken by ancestral NWMs as well as the paleoclimatic and paleogeographic events that may have affected them. Our analysis estimated a western range as the location of the most recent common ancestor of extant NWMs, this lead us to hypothesize about the possibility of an Asian origin of proto-platyrrhines.

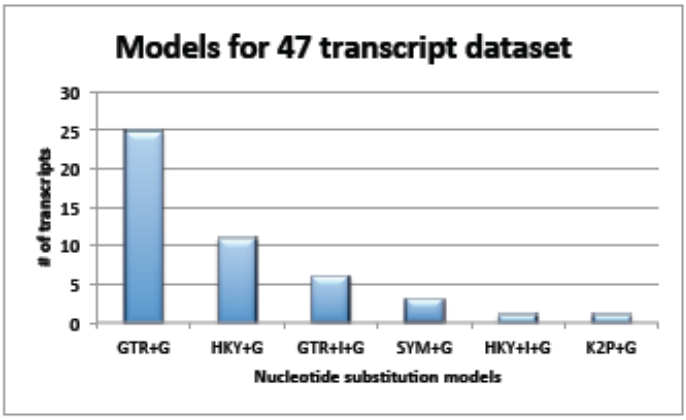
Finally, in Chapter 5, we attempt to connect changes in reproductive phenotype to concurrent genotypic changes among callitrichines. Our phylogenetic framework supports the placement of Goeldi's monkey nested within marmosets and tamarins suggesting the occurrence of two shifts in reproductive strategy (a gain of multiple offspring litters and a reversion back to singleton offspring) within approximately 10 million years. By analyzing genes involved in reproduction we have found that 12 genes show signs of adaptive evolution coincident with the changes in reproductive strategy. Among these genes are the leptin receptor (*LEPR*) and MMPs including *BMP8A*, *BMP8B*, and *BMPR1B*. Both *LEPR* and *BMPR1B* have been correlated with altered

ovulation number in other species. In addition three genes (*HAS1*, *LHB*, and *MMP9*) were identified as having sites undergoing accelerated evolution on the callitrichine lineages. In addition to considering genomic modifications associated with the mechanism of litter size alteration we also looked at physiological effects. We investigated a study population of Goeldi's monkeys for the presence of chimerism using both microsatellite genotyping and SRY amplification. In both instances all Goeldi's monkey individuals showed no sign of chimerism, thus ruling out late gestation twin resorption as the mechanism of litter size reduction. Our data suggest that litter size reduction in Goeldi's monkey occurs pre ovulation. The untested assumption of mono-ovulation in Goeldi's monkey is supported by our data, as many of the genes undergoing adaptive evolution coincident with the phenotypic changes are known to alter ovulation number. It is likely that the Goeldi's monkeys reverted back to mono-ovulation due to an increase in environmental stressors. Future studies on the timing and specific location of gene expression as well as functional studies of the genes found to be undergoing adaptive evolution will be necessary to provide additional insight into the reasons for the reversion to mono-ovulation.

APPENDIX A

CHAPTER 2 SUPPLEMENTAL MATERIAL

Figure S1: Optimal models of sequence evolution selected for each transcript in both datasets



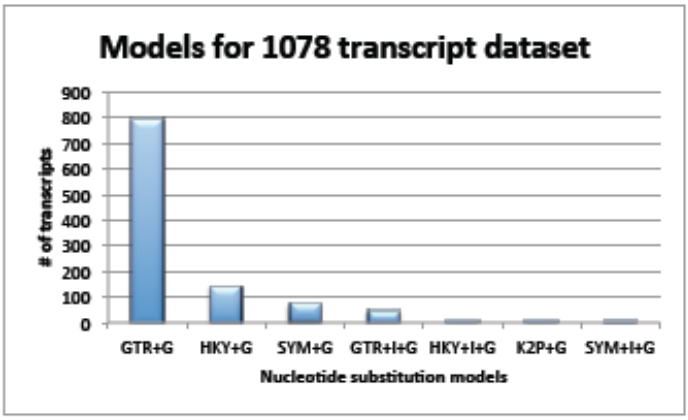


Table S1: Full list of taxa used in divergence dating analyses.

	Dataset source and size (bp)				Dataset source and size (bp)		
	1,078 transcript dataset	CFTR Region	CYP7A1 Region		1,078 transcript dataset	CFTR Region	CYP7A1 Region
Taxa	422,687	59,764	22,906	Taxa	422,687	59,764	22,906
Human	X	X	X	Galago	X	X	
Common Chimpanzee	X	X	X	Mouse Lemur	X	X	
Goriila		X		Ring tailed Lemur		X	X
Orangutan	X	X		Treeshrew	X		
Rhesus Monkey	X	X	X	Pika	X		
Baboon		X	X	Rabbit	X		
Vervet monkey		X		Mouse	X	X	
Black and white colobus			X	Rat	X	X	
Common marmoset	X	X	X	Kangaroo Rat	X		
Bolivian squirrel monkey		X	X	Guinea pig	X		
Owl monkey			X	Cow	X		
Dusky titi monkey		X		Dog	X	X	
Tarsier	X			Cat		X	

Table S2: Nucleotide composition in the 47 transcript and the 1,078 transcript datasets.

	47 transcript dataset				1078 transcript dataset			
Taxa	T (U)	C	A	G	T (U)	C	A	G
Human	22.2	24.7	26.9	26.2	24.6	22.6	28.8	24
Common chimpanzee	22.1	24.7	27	26.2	24.6	22.6	28.8	24
Orangutan	22.2	24.9	26.7	26.1	24.6	22.6	28.8	24
Rhesus monkey	22.3	24.6	27	26.2	24.7	22.6	28.8	24
Tarsier	22.2	24.8	26.9	26.1	24.6	22.6	28.8	24
Common marmoset	22.4	24.5	27	26.1	24.7	22.5	28.8	24
Mouse lemur	21.7	25.2	26.5	26.6	24.1	23.1	28.4	24.4
Galago	22.2	24.8	26.8	26.2	24.6	22.7	28.7	24
Treeshrew	21.7	25.2	26.8	26.4	23.8	23.6	28	24.6
Rabbit	20.8	26.3	25.7	27.2	23.3	24.1	27.4	25.3
Pika	20.6	26.4	25.4	27.6	22.9	24.5	26.9	25.7
Mouse	21.7	25.2	26.5	26.6	23.5	23.8	27.6	25.1
Rat	21.6	25.1	26.6	26.7	23.5	23.9	27.6	25.1
Kangaroo rat	21.7	25.3	26.6	26.4	24.3	23.1	28.4	24.2
Guinea pig	21.4	25.9	25.6	27.1	23.9	23.4	28	24.7
Cow	21.7	25.4	26.7	27.1	23.6	23.7	27.8	24.9
Dog	21.3	25.5	26.2	26.3	24.4	22.9	28.5	24.2
Average	21.7	25.2	26.5	26.5	24.1	23.2	28.2	24.5

Table S3: Divergence date estimates and 95% credibility intervals for all taxa using mcmctree.

	Bounded Fossil Calibrations			Probabilistic Calibrations		
Node	Mean	95% Credibility Interval		Mean	95% Credibility Interval	
Euarchontoglires - Laurasiatheria	83.3	78.1	91.9	82.9	77.8	91.5
Euarchonta - Glires	79.6	76.1	84.4	78.8	75.5	83.3
Scanadentia - Primates	78.2	74.9	82.9	77.4	74.2	81.8
Strepsirrhini - Haplorrhini	**72.6**	**69.6**	**76.9**	**71.8**	**68.9**	**75.8**
Anthropoidea - Tarsius	**68.6**	**65.6**	**72.7**	**67.8**	**64.9**	**71.7**
Catarrhini - Platyrrhini	40.0	37.3	43.1	38.0	33.8	42.0
Cercopithecoidea - Hominidae	25.4	23.7	27.6	24.3	21.8	26.9
Homininae - Pongina	18.0	16.8	19.6	16.1	13.6	18.7
Hominina - *Gorilla*	8.5	8.0	9.3	7.5	6.3	8.7
Homo - Pan	7.2	7.0	7.8	6.2	5.1	7.3
Cercopithecidae - Colobus	12.3	11.1	13.9	11.2	9.4	13.1
Chlorocebus - Macaca & Papio	9.0	8.2	10.1	8.2	7.0	9.6
Macaca - Papio	6.5	6.0	7.6	5.8	4.7	7.0
Cebidae - Pitheciidae	20.8	18.6	23.3	19.9	17.2	22.6
Cebinae & Aotinae - Callitrichinae	17.1	15.2	19.3	16.3	13.9	18.7
Cebinae - Aotinae	16.7	14.8	18.9	16.0	13.6	18.3
Microcebus & Lemur - Otolemur	52.4	47.0	57.2	51.2	45.5	56.3
Microcebus - Lemur	41.4	32.5	46.8	39.0	30.8	45.7
Lagomorpha - Rodentia	75.3	72.1	79.8	74.5	71.5	78.7
Oractolagus - Ochotona	53.7	53.0	56.2	53.6	53.0	56.0
Rodentia	65.6	64.0	68.8	65.3	63.8	68.0
Dipodomys-Muridae	59.7	57.3	62.9	59.5	57.1	62.2
Mus - Rattus	13.4	11.2	14.1	13.4	11.3	14.1
Bos - Carnivora	69.5	61.2	85.3	69.8	61.2	85.6
Felis - Canis	48.1	43.2	60.4	48.1	43.3	60.7

Table S4: Divergence date estimates and 95% credibility intervals for 17 mammalian taxa in BEAST using the 47 transcript dataset.

Divergence node	Mean (MYA)	95% Credibility Interval (MYA)	
Euarchontoglires - Laurasiatheria	91.2	71.4	119.9
Euarchonta - Glires	79.2	67.0	93.0
Scanadentia - Primates	74.7	62.2	89.4
Strepsirrhini - Haplorrhini	64.9	51.1	82.3
Anthropoidea - *Tarsius*	58.6	43.0	75.1
Catarrhini - Platyrrhini	38.6	33.7	43.7
Cercopithecidae - Hominidae	24.6	20.3	28.8
Hominina - Pongina	14.1	8.8	21.5
Homo - Pan	7.9	6.9	8.9
Microcebus - Otolemur	29.5	12.6	61.1
Lagomorpha - Rodentia	71.5	62.3	84.7
Oractolagus - Ochotona	56.7	51.8	61.8
Rodentia	55.2	50.5	60.1
Dipodomys-Muridae	27.7	20.6	36.7
Mus - Rattus	12.5	10.8	14.4
Bos - Canis	71.0	41.1	100.4

APPENDIX B

CHAPTER 3 SUPPLEMENTAL MATERIAL

Figure S2: Distribution of sequencing reads among human chromosomes based on BLAST search results

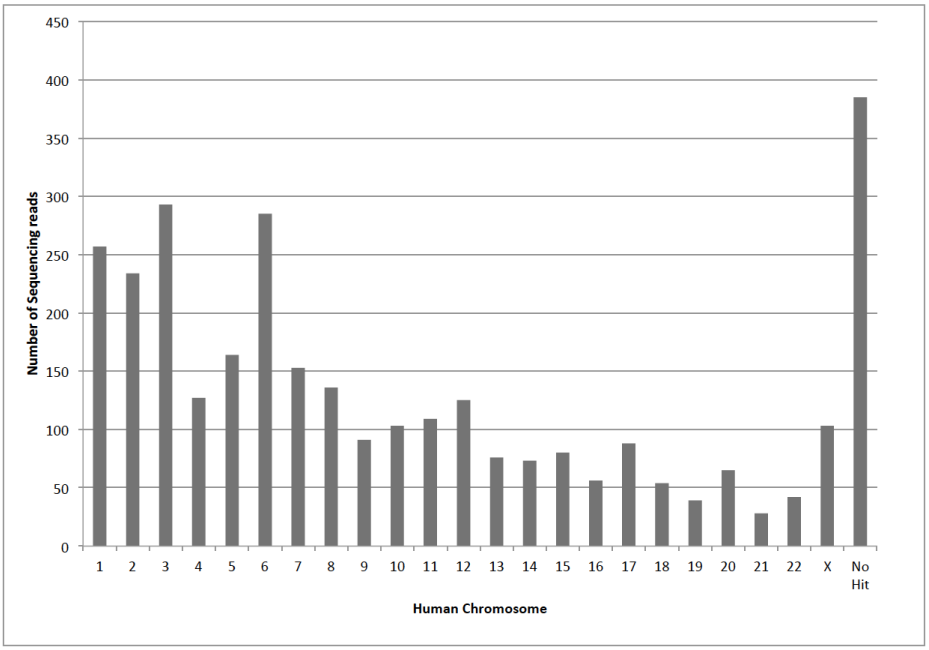


Table S5: Repetitive elements identified within sequencing reads

	Number of elements	Length occupied (bp)	Percentage of sequences
SINEs:	1629	308057	13.36%
ALUs	1266	260601	11.30%
MIRs	358	46834	2.03%
LINEs:	1039	330588	14.33%
LINE1	819	287262	12.45%
LINE2	193	38484	1.67%
L3/CR1	21	3789	0.16%
LTR element:	493	140229	6.08%
ERVL	109	28557	1.24%
ERVL-MaLRs	231	63644	2.76%
ERV_classI	129	44015	1.91%
ERV_classII	5	1378	0.06%
DNA elements:	292	56468	2.45%
hAT-Charlie	147	25150	1.09%
TcMar-Trigger	79	18638	0.81%
Unclassified:	9	1144	0.05%
Total interspersed repeats:		836486	36.27%
Small RNA:	31	2613	0.11%
Satellites:	235	48404	2.10%
Simple repeats:	284	15047	0.65%
Low complexity:	337	12963	0.56%

APPENDIX C

CHAPTER 4 SUPPLEMENTAL MATERIAL

Table S6: Primers for 64 non-genic markers

Marker	Primer sequence	Species used on
F_002	CAG CAT CTT AGA CCT GAC CAG TGA	All species
R_002	CCA GAT ACT CAT TGT GTC TCA CA	All species
F_003	GCC TAA GAT CTA ATC AGC ACA TTG	All species
R_003	TCT GTA TCT GCC TTT GTA AGA GG	All species
F_007	ATG ATG TCT CTG GTA GCA TA	All species
R_007	TCC ACC ATC TCA AGT AAT TTC AG	All species
F_008	AGG CTG ACC TAC CTT GAT TTC TTG	All species except below
F2_008	GTG AAA GGC TGA CCT ACC	All *Saguinus*, all *Callithrix* and *Callimico*
R_008	GCA TAT AAT TCC TCA TGC CTA GTG	All species
F_010	TCT CCT TCC CAT TCT GTG	All species
R_010	CAC ACC ACT TAG ATG GTC	All species
F_011	CAA TGG ATG GAT GAG AG	All species
R_011	GCA GAA CAC TCA GTG AGG TAC	All species
F_016	CCA CAC AGT CTA TGC TCT G	All species
R_016	GAG GCT GCT ATT CCT CAC	All species
F_020	CAT TCC AGA TGA AGT CCT C	All species
R_020	GTT TGA CAG TGA TGC CTT C	All species
F_029	GAC AGT TCT TTT GCA CCA G	All species
R_029	CCC ATT ACC TCT ATT AGT CTG C	All species
F_033	TGC CCT TCT CTA CTA AGT GG	All species
R_033	CCT TCC AGT TAA AGG TGC TC	All species
F_036	AGC CAC TAG CTT CAA GTG	All species
R_036	GTA GGT GGC TTG ATT GAG	All species
F_038	TAC CAC CTT CCT CTG GGT GTC	All species
R_038	TAG CCT ATC CCA CAG GCC AG	All species
F_043_044	AAG GAA CCA CCA CTA GGA G	All species
R_043_044	GTG CAA CAT TTA TCA TGA CTT C	All species
F_045	CCC TCA TAG TGC CTT CTG	All species
R_045	GAC CAG TAA AGG TTA GGG TC	All species
F_046	GTA AGG AAA TTA CAC TAC CTG TC	All species
R_046	GAT ACT GCT AGG ATG CCT G	All species
F_049_050	TCT GTC TGG GCA GAG TC	All species
R_049_050	GCA GTG ACA GTC CTA GTG	All species
F_055	CCA GTA GCG AGA GTT TGC	All species
R_055	GCC ATT CTT GCT GCT ACC	All species
F_061	CCA CTC TCA ATG CTG AGT AGG	All species
R_061	GAA CAT AGA CCG CTC CTG	All species
F_064	GGC ATT GAC AAA GTT ACA GGT GT	All species
R_064	GAG ACA TTA CCA GCA CAG AG	All species
F_065_066	CTA GAC CTG TGC AGT AGT C	All species
R_065_066	CTT CTC TGG CAC AGC CTG	All species
F2_080	GTG TCT TGA CAG GCT TTG	All species
R2_080	ACY TAC CTC ATG ATC TGC A	All species
F_085	GCT TCA AAC ACT CAG TGC TC	All species
R2_085	CAG GCA GYT TCA ATG ATT CTC	All species

Marker	Primer sequence	Species used on
216_F	GTG GAT AAG TTG TAG AAT GCT TC	All species
216_R	TTG TGC AAT AGC TGA TTA ACT C	All species
217_F	GTG AAC GGT GTT GAC AAG TGG	All species
217_R	CTC ATT GAC TGC CAT CTC AGC	All species
219_F	CAG ACA AGG CAG GCA GAT GA	All species
219_R	CTG ACC CTA TTC GTA TGC TAC C	All species
220_F	CTC TGT GCT GGG AGA CAA GG	All species
220_R	TGT GGC AAC AGA GTC AGG ATT C	All species
223_F	CCT CTC AAT GAG CAG AAA TAG TC	All species
223_R	CAA TTC AGT CAA GTT AGC CTC	All species
224_F	GAG AAA TTG CTG CTG ACT TG	All species
224_R	CCT GGA GAT GCA ATC ATG TTG	All species
232_F	GTC TCT GCT GAG AAC TGA GGC	All species
232_R	AGG GCA GGC AGG TAG TGG	All species
250_F	CAA TGC AAT ATA TAC CTC C	All species
250_R	GGG TCT GCT GAA GAA GTT AAT G	All species
252_F	GAG CAG TCA TGA TAC AAA ATA GGG	All species
252_R	CTC AAG TAT TCT TTG TAT CTG GC	All species
254_F	ACA TAT GAC TCA GGC CAA ATC	All species
254_R	GCA TTG CAG TAG CTA GCA C	All species
258_F	CCA AGG CAT AGT GTC TTA ACA	All species
258_R	AAC CTG TCC CTG TAT CTA AAA C	All species
263 F	ACA AGT TAG CAT CTG ATT CAT TTA C	All species
263 R	CAA CTT TCC TGG TAG CTT TGA	All species
265 F	TCC ATA GTA CAC CAA GGG ACC	All species
265 R	CCA ATA GTA TGC ACT GTG AGC ATG	All species
266 F	CCA CTA AGG AAT TCT GAT GCT T	All species
266 R	GCA TCA CCC ATG TAT GAT ACA TC	All species
267 F	GTA ACA ATG CTA GAC AGG ATT G	All species
267 R	GAG TAA TTC ACC TTG CTG AGA C	All species
271_F	CTT GAA CAA CAC ATA TTT CCA ACC	All species
271_R	GAA AGA TGG CAT CTA CTG GTG A	All species
34_T3_F	GTA AAG TAT CCT TAT TGC AGC ATG	All species
34_T3_R	CCA ACT TAA TAC CTA TTT TGT CTG	All species
64_T3_F	CAC TAT TTC TAG AAA GAG CTG GCC CAC	All species
64_T3_R	TAT CCT AAT TCT GCA TCT CTG AAC ATG C	All species
B22_F	CAT GAA GCA CCC CCA GAA CAG TCA G	All species
B22_R	TGG ATG AGA TGG ATA CAA AAG AAC TAC AAC TC	All species
C11_02_F	CTG GTC AAG AAC ACT ATC CAC	All species
C11_02_R	GCC TGA AGA TAT ACT CAA TGC C	All species
C11_04_F	GAC TTA GAG CCC ACT CTG CA	All species
C11_04_R	GCT TAC AGA TTT CAG TTA AGC CAA	All species
C12_01_F	GAC TAA ATT AGG AAT TGG ATA ATA TGG G	All species
C12_01_R	TGT ACC ACT ATT AAA TGT TGT GAT GG	All species

Table S6: Primers for 64 non-genic markers continued

Marker	Primer sequence	Species used on	Marker	Primer sequence	Species used on
F_093	TAA CTA GAG TGA ACA TTT GGA CTG	All species	C13_02_F	GCC ACT TGA GAA ATC CCA TTT G	All species
R_093	ATA GAG GGC TAC AGA ACG	All species	C13_02_R	GCT GAC CAT GTA GTG AAG TC	All species
105_T3_F	GCT GAT TTC ACC TGT TTC ACA GTG	All species	C13_04_F	CAG CCC AGA GTG CTT AGT TC	All species
105_T3_R3	ACC AAG TGC TGA GGA GCT TAG	All species	C13_04_R	GAA GCA GAT AGG CAG CCA C	All species
113_T7_F2	CAT GCT AGC AAG TAA GCT TGT C	All species	C16_01_F	CCA CAT TTC TGG GGA AGC TTC	All species
113_T7_R2	GAG ATG GGT CCT TCC TGT CC	All species	C16_01_R	GGA AAG CAT TAC AAA GGT GGA ACC	All species
190_F2	CAC TGG CCA TYC AGC CTC CTG GT	All species	C16_02_F	GGA GAC AGG AGG TAT GAG A	All species
190_R2	TGT CTG GTG CTA CAT CCA GAG C	All species	C16_02_R	GAG CTG AGT GCA GGC GCT	All species
194_F	ATT CTA TTC CCT GTG ATG AWA GCA GA	All species	C17_01F	GTC CTG GAT TTC CTA TCA CC	All species
194_R	TCT TTT CAC TCA ACA TAT GCC TGG A	All species	C17_01R	GGA GCT GTC TTC CTC TGT AA	All species
201_F	GGT AAC CAG TTT GGG CTT CTG	All species	C17_04F	CCT ACT GGT CTC AGG GCA T	All species
201_R	CAT CGT CCT GGT TGT GAA GTG	All species	C17_04R	TTG TGG TCT GGA CAG GCC	All species
202_F	GCC TTG ATA GTG TGA AGT TAT TCC	All species	C18_03F	GCA TAC ATG CTT GCA TCC AA	All species
202_R	TCA CTA AAG ACA GCT CTT CAT CAC	All species	C18_03R	GCA TAG TAG GGG TGC ACA G	All species
21_T3_F	GCC CTT YGA GTA CTT CCC ATA TGC	All species	C19_03F	CCT TCC CAC TGG TAA ATG AAT C	All species
21_T3_R	GAC CCA AGT CAT AAC CAA TGG	All species	C19_03R	GAG ATG CAT GCT TCT CCT G	All species
210_F	CAG GTA GAG CTG CCT CTT CA	All species	C21_01F	GTC CAG TTG CTT ATT AAG AGC AC	All species
210_R	ACT TCA ATG GTA CAC TTT CGT T	All species	C21_01R	GAT GAC TAA TTG ATG TCA ACA GGT	All species
215_F	CAA AGA TGA GGC AAA GCA TAG AT	All species	E22_F2	AGA AGT AAA AGA AAT KTT ATT AAC ACC TG	All species
215_R	GCA ATC CGT ATT TAA GGT GCT AG	All species	E22_R2	CTA GMC TCT TAA CTA AAT TGT GCT A	All species

Table S7: Division of partitions based on GC content

Upper Outliers		75 to 95		Mid to 75	
Markers included	Averaged GC	Markers included	Averaged GC	Markers included	Averaged GC
C19_03	51.92%	21_T3	42.61%	266	39.80%
C21_01	58.78%	210	45.77%	267	38.84%
E22	48.94%	215	43.88%	271	39.55%
		216	47.54%	34_T3	41.41%
		217	46.72%	64_T3	38.46%
		219	43.90%	B22	42.49%
		220	44.56%	C11_02	39.08%
		223	45.43%	C11_04	42.50%
		224	43.96%	C12_01	42.37%
		232	46.26%	C13_02	42.43%
		250	46.34%	C13_04	39.71%
		252	46.17%	C16_01	41.80%
		254	46.88%	C16_02	40.23%
				C17_01	38.85%
				C17_04	41.88%
				C18_03	38.51%
Lower Outliers		**5 to 25**		**25 to Mid**	
Markers included	Averaged GC	Markers included	Averaged GC	Markers included	Averaged GC
258	27.59%	055	29.52%	002	37.82%
263	27.66%	061	31.46%	003	38.20%
265	26.81%	064	30.56%	007	38.12%
		065_066	31.55%	008	33.01%
		080	32.08%	010	35.97%
		085	31.71%	011	33.36%
		093	32.51%	016	33.86%
		105_T3	29.68%	020	37.03%
		113_T3	31.34%	029	38.16%
		190	30.68%	033	33.52%
		194	29.31%	036	35.95%
		201	30.81%	038	35.54%
			30.71%	043_044	37.12%
				045	35.40%
				046	38.02%
				049_050	33.20%

Table S8: Annotation of non-genic markers

Marker ID	Alignment Length	GC Content	Taxa Coverage	Human Chromosome #	Human Chromosome position	Marmoset Chromosome #	Marmoset Chromosome position	Model of sequence evolution
002	322	42.61%	100.00%	1	62848172 62848478	7	97685869 97686185	GTR+G
003	989	29.52%	100.00%	1	80453685 80454162	7	115837969 115838433	GTR
007	276	31.46%	87.50%	1	88854999 88855273	7	124293564 124293831	GTR
008	476	39.80%	97.50%	1	221308924 221309395	19	3489301 3489764	GTR+I
010	286	51.92%	77.50%	2	67232836 67233121	14_GL285651_random	132654 132939	K80+G
011	241	38.84%	90.00%	2	67673593 67674012	14	39497707 39498114	GTR+G
016	563	30.56%	97.50%	2	154021856 154022415	6	66148563 66149119	GTR+G
020	563	39.55%	95.00%	3	59472363 59472920	15	35550370 35550917	HKY+G
029	499	37.82%	87.50%	3	157630390 157630888	17	19450797 19451283	HKY+G
033	369	45.77%	90.00%	4	58364461 58364827	3	136000342 136000706	HKY+G
036	538	41.41%	100.00%	5	55697761 55698295	2	151245327 151245860	HKY+I
038	708	38.20%	100.00%	5	71366492 71367192	2	136501373 136502072	GTR+G
043_044	951	38.12%	100.00%	6	76786479 76787373	4	74422203 74423087	GTR+G
045	329	33.01%	90.00%	7	93742822 93743135	8	53397124 53397437	HKY+G
046	545	35.97%	72.50%	7	115501713 115502240	8	84731090 84731630	HKY+G
049_050	692	43.88%	92.50%	7	141029112 141029794	8	104750110 104750786	HKY+G
055	479	38.46%	87.50%	8	31289156 31289620	13	29438890 29439353	GTR+G
061	533	47.54%	92.50%	9	108438812 108439337	1	149018915 149019442	GTR+G
064	576	42.49%	97.50%	X	151042661 151043137	X	138816279 138816751	HKY
065_066	588	39.08%	95.00%	10	60217767 60218344	12	65928193 65928773	GTR+G
080	449	33.36%	100.00%	15	46026592 46027040	10	22718780 22719228	GTR+G
085	381	33.86%	95.00%	20	52483875 52484231	5	47755572 47755933	HKY+I
093	461	42.50%	85.00%	14	46837282 46837730	10	70898147 70898589	HKY+I
105_T3	478	46.72%	95.00%	6	20312930 20313405	4	8226763 8227239	GTR+G
113_T3	1258	42.37%	90.00%	2	36518317 36518970	14	72026266 72026916	GTR+I
190	641	58.78%	82.50%	5	2417788 2418417	2	202134045 202134666	GTR+G
194	472	37.03%	95.00%	10	19829175 19829640	7	17253985 17254448	GTR
201	747	31.55%	100.00%	5	92144450 92145177	2	115046746 115047470	GTR+G
202	404	38.16%	92.50%	5	102628827 102629221	2	104107930 104108325	HKY+G
21_T3	648	42.43%	92.50%	X	99732138 99732780	X	89716761 89717403	HKY+G
210	690	32.08%	95.00%	5	106692479 106693149	2	99832026 99832706	HKY+G
215	487	33.52%	90.00%	5	124124813 124125287	2	81254803 81255276	GTR+G
216	424	27.59%	90.00%	5	119013117 119013538	2	86761469 86761890	GTR
217	547	39.71%	95.00%	5	81131661 81132207	2	126323406 126323952	GTR+I
219	624	43.90%	95.00%	5	67633004 67633624	2	138724435 138725053	GTR+G
220	533	41.80%	95.00%	5	61996689 61997152	2	144664230 144664681	GTR+G
223	651	35.95%	95.00%	5	53608397 53608985	2	153387620 153388209	HKY+G

Table S8: Annotation of non-genic markers continued

Marker ID	Alignment Length	GC Content	Taxa Coverage	Human Chromosome #	Human Chromosome position	Marmoset Chromosome #	Marmoset Chromosome position	Model of sequence evolution
224	1086	40.23%	95.00%	5	6945093 6946138	2	19722176219722280 1	GTR+G
232	1047	44.56%	95.00%	5	13609702 13610424	2	19000818819000889 6	GTR+G
250	471	45.43%	95.00%	3	135556308 135556775	17	40671271 40671734	HKY+G
252	767	31.71%	97.50%	5	29234093 29234841	2	17409440817409515 4	GTR+G
254	464	32.51%	85.00%	5	29197959 29198391	2	17413426517413470 9	GTR+G
258	727	27.66%	92.50%	5	29138274 29138992	2	17419686717419758 7	GTR+G
263	562	26.81%	87.50%	5	29014630 29015158	2	17432047517432101 2	GTR+G
265	847	29.68%	92.50%	5	28988346 28989178	2	17434179717434263 0	GTR+G
266	815	31.34%	85.00%	5	28979126 28979875	2	17434868117434933 2	GTR+G
267	1061	30.68%	85.00%	5	28957777 28958771	2	17437203817437302 0	GTR+G
271	708	29.31%	85.00%	5	28903250 28903947	2	17442656117442725 6	GTR+G
34_T3	892	35.54%	90.00%	1	74070379 74070969	7	10945010710945069 7	HKY
64_T3	652	48.94%	65.00%	10	86907408 86908018	12	38971723 38972353	K80+G
B22	689	43.96%	87.50%	3	71721027 71721701	15	48027415 48028093	GTR+G
C11_02	679	37.12%	67.50%	11	7970858 7971524	11	71342657 71343323	HKY+G
C11_04	718	35.40%	70.00%	11	20219123 20219831	11	83570817 83571517	GTR+G
C12_01	578	30.81%	82.50%	12	30940336 30940876	9	36461657 36462191	HKY
C13_02	816	38.85%	95.00%	13	28283517 28284301	5	140682241 140683025	HKY+G
C13_04	693	38.02%	90.00%	13	59646675 59647341	1	55969463 55970107	GTR+G
C16_01	629	46.26%	90.00%	16	52743353 52743978	20	6590485 6591125	HKY+G
C16_02	683	46.34%	95.00%	16	72254572 72255248	20	26086886 26087544	GTR+G
C17_01	613	46.17%	80.00%	17	13853061 13853657	5	67618782 67619387	HKY+G
C17_04	1180	41.88%	92.50%	17	53311311 53312093	5	102094309 102095092	GTR+G
C18_03	587	33.20%	82.50%	18	24233876 24234450	13	64237741 64238313	GTR+G
C19_03	728	46.88%	75.00%	19	31469564 31470285	22	25596179 25596873	GTR+G
C21_01	500	30.71%	100.00%	21	16630899 16631628	21	19966467 19967209	HKY
E22	656	38.51%	92.50%	1	198080361 198081003	19	19557985 19558633	GTR+G

APPENDIX D

CHAPTER 5 SUPPLEMENTAL MATERIAL

Table S9: Species coverage for each gene analyzed in PAML

Gene ID	Transcript ID	Gene	Human	Chimpanzee	Gorilla	Orangutan	Macaque	Titi monkey	Spider monkey	Owl monkey	Goeldi's monkey	Marmoset	Tarsier	Galago	Mouse lemur	Guinea pig	Mouse	Rat
ENSG00000115170	ENST00000263640	*ACVR1*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000121989	ENST00000241416	*ACVR2A*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000114739	ENST00000352511	*ACVR2B*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000135744	ENST00000366667	*AGT*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000116127	ENST00000264448	*ALMS1*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000095739	ENST00000375533	*BAMBI*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000168487	ENST00000306385	*BMP1*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000125845	ENST00000378827	*BMP2*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000138756	ENST00000335016	*BMP2K*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000125378	ENST00000245451	*BMP4*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000112175	ENST00000370830	*BMP5*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000153162	ENST00000283147	*BMP6*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000101144	ENST00000395863	*BMP7*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000183682	ENST00000331593	*BMP8A*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000116985	ENST00000372827	*BMP8B*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000107779	ENST00000224764	*BMPR1A*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000138696	ENST00000515059	*BMPR1B*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000204217	ENST00000374580	*BMPR2*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000118971	ENST00000261254	*CCND2*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000172216	ENST00000303004	*CEBPB*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000135346	ENST00000369582	*CGA*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000132693	ENST00000255030	*CRP*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000136488	ENST00000316193	*CSH1*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000213218	ENST00000392886	*CSH2*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000204414	ENST00000309894	*CSHL1*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000135047	ENST00000343150	*CTSL1*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000120738	ENST00000239938	*EGR1*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000140009	ENST00000341099	*ESR2*	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x
ENSG00000118689	ENST00000406360	*FOXO3*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

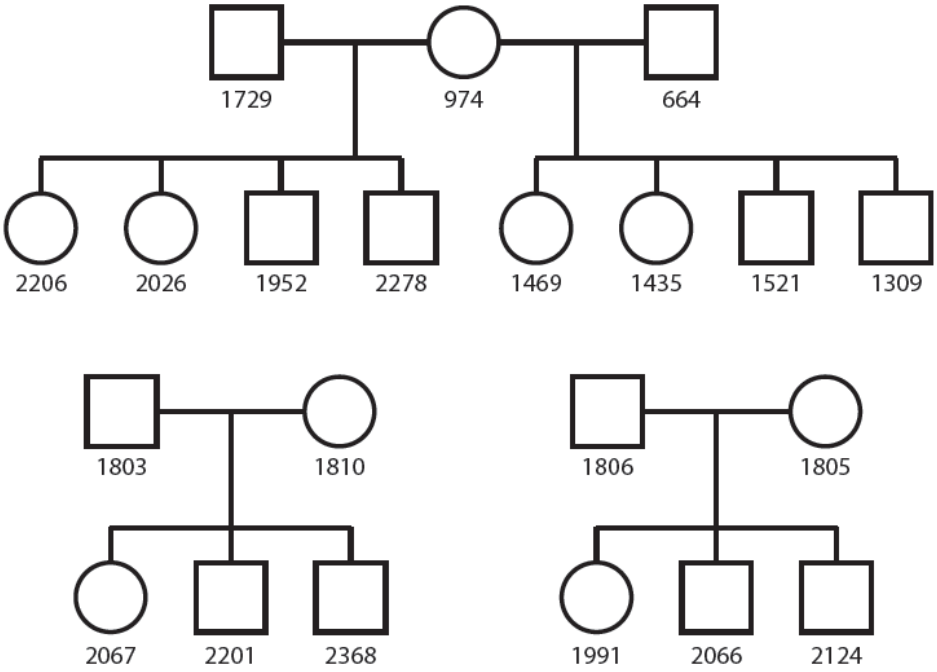
Table S9: Species coverage for each gene analyzed in PAML, Continued

Gene ID	Transcript ID	Gene	Human	Chimpanzee	Gorilla	Orangutan	Macaque	Titi monkey	Spider monkey	Owl monkey	Goeldi's monkey	Marmoset	Tarsier	Galago	Mouse lemur	Guinea pig	Mouse	Rat
ENSG00000134363	ENST00000256759	*FST*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000163430	ENST00000295633	*FSTL1*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000053108	ENST00000265342	*FSTL4*	×	×	×	×	×			×	×	×	×	×	×	×	×	×
ENSG00000135414	ENST00000257868	*GDF11*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000164404	ENST00000378673	*GDF9*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000259384	ENST00000323322	*GH1*	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×
ENSG00000136487	ENST00000332800	*GH2*	×	×	×	×	×	×		×	×	×	×	×	×		×	×
ENSG00000112964	ENST00000230882	*GHR*	×	×	×	×	×	×	×	×	×	×		×	×	×	×	×
ENSG00000105509	ENST00000222115	*HAS1*	×	×	×	×	×	×			×	×	×	×	×	×	×	×
ENSG00000170961	ENST00000303924	*HAS2*	×	×		×	×	×	×		×	×	×	×	×	×	×	×
ENSG00000100644	ENST00000337138	*HIF1A*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000117594	ENST00000261465	*HSD11B1*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000017427	ENST00000307046	*IGF1*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000140443	ENST00000268035	*IGF1R*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000159217	ENST00000290341	*IGF2BP1*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000073792	ENST00000382199	*IGF2BP2*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000136231	ENST00000258729	*IGF2BP3*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000197081	ENST00000356956	*IGF2R*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000115457	ENST00000233809	*IGFBP2*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000141753	ENST00000269593	*IGFBP4*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000163453	ENST00000295666	*IGFBP7*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000115008	ENST00000263339	*IL1A*	×	×	×	×	×				×	×	×	×	×	×	×	×
ENSG00000125538	ENST00000263341	*IL1B*	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×
ENSG00000077238	ENST00000395762	*IL4R*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000122641	ENST00000242208	*INHBA*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000151655	ENST00000358415	*ITIH2*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000116678	ENST00000349533	*LEPR*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000100097	ENST00000215909	*LGALS1*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000131981	ENST00000254301	*LGALS3*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ENSG00000168961	ENST00000395473	*LGALS9*	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×

Table S9: Species coverage for each gene analyzed in PAML, Continued

Gene ID	Transcript ID	Gene	Human	Chimpanzee	Gorilla	Orangutan	Macaque	Titi monkey	Spider monkey	Owl monkey	Goeldi's monkey	Marmoset	Tarsier	Galago	Mouse lemur	Guinea pig	Mouse	Rat
ENSG00000104826	ENST00000221421	*LHB*	x	x	x	x	x	x			x	x		x	x	x	x	x
ENSG00000128342	ENST00000249075	*LIF*	x	x	x	x	x		x		x	x	x	x	x	x	x	x
ENSG00000135679	ENST00000462284	*MDM2*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000198625	ENST00000367182	*MDM4*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000106484	ENST00000223215	*MEST*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000013774	ENST00000260302	*MMP13*	x	x	x	x	x				x	x	x	x	x	x	x	x
ENSG00000157227	ENST00000311852	*MMP14*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000123342	ENST00000322569	*MMP19*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000087245	ENST00000219070	*MMP2*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000100985	ENST00000372330	*MMP9*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000177000	ENST00000376592	*MTHFR*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000198793	ENST00000361445	*mTOR*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000180530	ENST00000400202	*NRIP1*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000082175	ENST00000325455	*PGR*	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x
ENSG00000151665	ENST00000281382	*PIGF*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000104368	ENST00000220809	*PLAT*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000132170	ENST00000287820	*PPARG*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000073756	ENST00000367468	*PTGS2*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000096717	ENST00000212015	*SIRT1*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000117394	ENST00000426263	*SLC2A1*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000111371	ENST00000439706	*SLC38A1*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000157637	ENST00000374759	*SLC38A10*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000134294	ENST00000256689	*SLC38A2*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000139209	ENST00000447411	*SLC38A4*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000017483	ENST00000376876	*SLC38A5*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000139974	ENST00000354886	*SLC38A6*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000103042	ENST00000570101	*SLC38A7*	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000177058	ENST00000396865	*SLC38A9*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000102265	ENST00000218388	*TIMP1*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000035862	ENST00000262768	*TIMP2*	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
ENSG00000100234	ENST00000266085	*TIMP3*	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
ENSG00000232810	ENST00000449264	*TNF*	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x
ENSG00000141510	ENST00000269305	*TP53*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000187555	ENST00000344836	*USP7*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000112715	ENST00000372055	*VEGFA*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000173511	ENST00000309422	*VEGFB*	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
ENSG00000105663	ENST00000222270	*WBP7*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ENSG00000127578	ENST00000319070	*WFIKKN1*	x		x	x	x		x	x	x	x	x	x	x	x	x	x

Figure S3: Pedigree of *Callimico goeldii* individuals tested for chimerism. Tissue from all individuals was obtained from the Brookfield zoo in Chicago Il, numbers indicate the individuals studbook ID



REFERENCES

Altmann, J., Warneke, M., Ramer, J., 1988. Twinning among Callimico goeldii. International Journal of Primatology 9, 165-168.

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.

Andrews, T.D., Jermiin, L.S., Easteal, S., 1998. Accelerated evolution of cytochrome b in simian primates: adaptive evolution in concert with other mitochondrial proteins? Journal of molecular evolution 47, 249-257.

Arnason, U., Adegoke, J.A., Gullberg, A., Harley, E.H., Janke, A., Kullberg, M., 2008. Mitogenomic relationships of placental mammals and molecular estimates of their divergences. Gene 421, 37-51.

Asher, R.J., Meng, J., Wible, J.R., McKenna, M.C., Rougier, G.W., Dashzeveg, D., Novacek, M.J., 2005. Stem Lagomorpha and the antiquity of Glires. Science (New York, N.Y.) 307, 1091-1094.

Baba, M.L., Weiss, M.L., Goodman, M., Czelusniak, J., 1982. The Case of Tarsier Hemoglobin. Systematic Zoology 31, 156-165.

Babb, P.L., Fernandez-Duque, E., Schurr, T.G., 2010. *AVPR1A* sequence variation in monogamous owl monkeys (*Aotus azarai*) and its implications for the evolution of platyrrhine social behavior. Journal of molecular evolution 71, 279-297.

Bailey, W.J., Fitch, D.H., Tagle, D.A., Czelusniak, J., Slightom, J.L., Goodman, M., 1991. Molecular evolution of the psi eta-globin gene locus: gibbon phylogeny and the hominoid slowdown. Mol Biol Evol 8, 155-184.

Bailey, W.J., Hayasaka, K., Skinner, C.G., Kehoe, S., Sieu, L.C., Slightom, J.L., Goodman, M., 1992. Reexamination of the African hominoid trichotomy with additional sequences from the primate beta-globin gene cluster. Mol Phylogenet Evol 1, 97-135.

Barroso, C.M.L., Schneider, H., Schneider, M.P.C., Sampaio, I., Harada, M.L., Czelusniak, J., Goodman, M., 1997. Update on the phylogenetic systematics of new world monkeys: Further DNA evidence for placing the pygmy marmoset (Cebuella) within the genus Callithrix. International Journal of Primatology 18, 651-674.

Beard, K.C., 1998a. East of Eden: Asia as an important center of taxonomic origination in mammalian evolution. Bull. Carnegie Mus Nat Hist 34, 5-39.

Beard, K.C., 1998b. A new genus of Tarsiidae Mammalia: Primates from the middle eocene of Shanxi Province, China, with notes on the historical biogeography of tarsiers. Bulletin of Carnigie Museum of Natural History 34, 260-277.

Beard, K.C., Qi, T., Dawson, M.R., Wang, B., Li, C., 1994. A diverse new primate fauna from middle Eocene fissure-fillings in southeastern China. Nature 368, 604-609.

Benirschke, K., 1995. The biology of the twinning process: how placentation influences outcome. Semin Perinatol 19, 342-350.

Benirschke, K., Anderson, J.M., Brownhill, L.E., 1962. Marrow Chimerism in Marmosets. Science (New York, N.Y.) 138, 513-515.

Berggren, W.A., Hollister, C.D., 1974. Paleogeography, Paleobiography and the history of circulation in the Atlantic Ocean. In: Hay, W.W. (Ed.), Studies in Paleo-Oceanography. Society of Economic Paleontologists and Mineralogists, Tulsa, pp. 126-186.

Bindon, B.M., 1984. Reproductive biology of the Booroola Merino sheep. Australian journal of biological sciences 37, 163-189.

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. Proc Natl Acad Sci U S A 104, 1159-1164.

Bonner, T.I., Heinemann, R., Todaro, G.J., 1980. Evolution of DNA sequences has been retarded in Malagasy primates. Nature 286, 420-423.

Brandley, M.C., Schmitz, A., Reeder, T.W., 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54, 373-390.

Cachel, S., 1981. Plate tectonics and the problem of anthropoid origins. Yearb. Phys. Anthropol. 24, 139-172.

Campbell, B.K., Baird, D.T., Souza, C.J., Webb, R., 2003. The FecB (Booroola) gene acts at the ovary: in vivo evidence. Reproduction (Cambridge, England) 126, 101-111.

Campfield, L.A., Smith, F.J., Guisez, Y., Devos, R., Burn, P., 1995. Recombinant mouse OB protein: evidence for a peripheral signal linking adiposity and central neural networks. Science (New York, N.Y.) 269, 546-549.

Canavez, F., Alves, G., Fanning, T.G., Seuanez, H.N., 1996. Comparative karyology and evolution of the Amazonian Callithrix (Platyrrhini, Primates). Chromosoma 104, 348-357.

Canavez, F.C., Moreira, M.A., Ladasky, J.J., Pissinatti, A., Parham, P., Seuanez, H.N., 1999. Molecular phylogeny of new world primates (Platyrrhini) based on beta2-microglobulin DNA sequences. Mol Phylogenet Evol 12, 74-82.

Carroll, J., 1982. Maintenance of the Goeldi's monkey, *Callimico goeldii,* at Jersey Wildlife Preservation trust. Int Zoo Yearbk 22, 101-105.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17, 540-552.

Chatterjee, H.J., Ho, S.Y., Barnes, I., Groves, C., 2009. Estimating the phylogeny and divergence times of primates using a supermatrix approach. BMC evolutionary biology 9, 259.

Chaves, R., Sampaio, I., Schneider, M.P., Schneider, H., Page, S.L., Goodman, M., 1999. The place of *Callimico Goeldii* in the callitrichine phylogenetic tree: Evidence from von Willebrand factor gene intron II sequences. Molecular Phylogenetics and Evolution 13, 392-404.

Chehab, F.F., Lim, M.E., Lu, R., 1996. Correction of the sterility defect in homozygous obese female mice by treatment with the human recombinant leptin. Nature genetics 12, 318-320.

Chehab, F.F., Mounzih, K., Lu, R., Lim, M.E., 1997. Early onset of reproductive function in normal female mice treated with leptin. Science (New York, N.Y.) 275, 88-90.

Chen, F.C., Li, W.H., 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am J Hum Genet 68, 444-456.

Ciochon, R.L., Chiarelli, A.B., 1980a. Evolutionary Biology of the New World Monkeys and Continental Drift. Plenum Publishing Corporation, New York NY.

Ciochon, R.L., Chiarelli, A.B., 1980b. Paleobiogeographic perspectives on the origin of the Platyrrhini. In: Ciochon, R.L., Chiarelli, A.B. (Eds.), Evolutionary Biology of the New World monkey and Continental Drift. Plenum, New York, pp. 459-494.

Clisson, I., Lathuilliere, M., Crouau-Roy, B., 2000. Conservation and evolution of microsatellite loci in primate taxa. Am J Primatol 50, 205-214.

Coleman, D.L., 1973. Effects of parabiosis of obese with diabetes and normal mice. Diabetologia 9, 294-298.

Comejo, F., 2008. Callimico goeldii. IUCN Red List of Threatened Species Version 2013.1.

Consortium, T.C.S.a.A., 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437, 69-87.

Cooper, G.M., Stone, E.A., Asimenos, G., Green, E.D., Batzoglou, S., Sidow, A., 2005. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 15, 901-913.

Crisci, J.V., 2001. The voice of historical biogeography. Journal of Biogeography 28, 157-168.

Cronin, J.E., Sarich, V.M., 1978. Marmoset evolution: the molecular evidence. Primates Med 10, 12-19.

Davis, G.H., 2005. Major genes affecting ovulation rate in sheep. Genetics, selection, evolution : GSE 37 Suppl 1, S11-23.

de Jong, W.W., Goodman, M., 1988. Anthropoid affinities of *Tarsius* supported by lens [alpha]A-crystallin sequences. Journal of human evolution 17, 575-582.

Decker, J.E., Pires, J.C., Conant, G.C., McKay, S.D., Heaton, M.P., Chen, K., Cooper, A., Vilkki, J., Seabury, C.M., Caetano, A.R., Johnson, G.S., Brenneman, R.A., Hanotte, O., Eggert, L.S., Wiener, P., Kim, J.J., Kim, K.S., Sonstegard, T.S., Van

Tassell, C.P., Neibergs, H.L., McEwan, J.C., Brauning, R., Coutinho, L.L., Babar, M.E., Wilson, G.A., McClure, M.C., Rolf, M.M., Kim, J., Schnabel, R.D., Taylor, J.F., 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. Proc Natl Acad Sci U S A 106, 18644-18649.

Delson, E., 1992. Evolution of Old World Monkeys. In: Jones, S., Martin, R.D., Pilbeam, D.R. (Eds.), Cambridge Encyclopedia of Human Evolution. Press Syndicate of the University of Cambridge, Cambridge, pp. 217-222.

Delson, E., Rosenberger, A.L., 1980. Phyletic perspectives on platyrrhine origins and anthropoid relationships. In: Ciochon, R.L., Chiarelli, A.B. (Eds.), Evolutionary Biology of the New World monkeys and continental drift. Plenum, New York, pp. 445-458.

Delson, E., Tattersall, I., Van Couvering, J.A., Brooks, A.S., 2000. Cercopithecidae. In: Delson, E., Tattersall, I., Van Couvering, J.A., Brooks, A.S. (Eds.), Encyclopedia of Human Evolution and Prehistory, 2nd Ed. Garland, New York, pp. 166-171.

Dixon, A.F., 1983. The Owl monkey (*Aotus trivirgatus*). In: Hearn, J.P. (Ed.), Reproduction in New World Primates-New Models in Medical Science. MTP Press Limited, London.

Djian, P., Green, H., 1991. Involucrin gene of tarsioids and other primates: alternatives in evolution of the segment of repeats. Proc Natl Acad Sci U S A 88, 5321-5325.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969-1973.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.

Eisenberg, J., 1981. The mammalian radiations. University of Chicago Press, Chicago.

Elango, N., Thomas, J.W., Yi, S.V., 2006. Variable molecular clocks in hominoids. Proc Natl Acad Sci U S A 103, 1370-1375.

Ellegren, H., Smith, N.G., Webster, M.T., 2003. Mutation rate variation in the mammalian genome. Current opinion in genetics & development 13, 562-568.

Elliot, M.G., Crespi, B.J., 2009. Phylogenetic evidence for early hemochorial placentation in eutheria. Placenta 30, 949-967.

Enders, A.C., Lopata, A., 1999. Implantation in the marmoset monkey: expansion of the early implantation site. Anat Rec 256, 279-299.

Escobar-Paramo, P., 2000. Microsatellite primers for the wild brown capuchin monkey *Cebus apella*. Molecular Ecology 9, 107-108.

Fabre, P.H., Rodrigues, A., Douzery, E.J.P., 2009. Patterns of macroevolution among Primates inferred from a supermatrix of mitochondrial and nuclear DNA. Molecular Phylogenetics and Evolution 53, 808-825.

Felsenstein, J., 1978. Cases in which Parsimony or Compatibility Methods will be Positively Misleading. Syst Biol 27, 401-410.

Felsenstein, J., 2004. Inferring Phylogenies. Sinauer Associates, Sunderland MA.

Ferrari, S.F., 1993. Ecological differentiation in the Callithrichidae. In: Rylands, A.B. (Ed.), Marmosets and Tamarins: Systematics, Ecology, and Behaviour. Oxford University Press, Oxford.

Ferrari, S.F., Lopes Ferrari, M.A., 1989. A re-evaluation of the social organisation of the Callitrichidae, with reference to the ecological differences between genera. Folia Primatol (Basel) 52, 132-147.

Fleagle, J.G., Tejedor, M.F., 2002. Early platyrrhines of southern South America. In: Hartwig, W.C. (Ed.), The Primate Fossil Record. Cambridge University Press, Cambridge.

Ford, S.M., 1980. Callitrichids as phyletic dwarfs, and the place of the callitrichidae in platyrrhini. Primates 21, 31-43.

Ford, S.M., 1986. Systematics of the New World monkeys. In: Swindler, D.R., Erwin, J. (Ed.), Comparative Primate biology Alan Liss, New York, pp. 73-135.

Ford, S.M., Davis, L.C., 1992. Systematics and body size: Implications for feeding adaptations in new world monkeys. American journal of physical anthropology 88, 415-468.

Fragaszy, D.M., Adams-Curtis, L.E., 1998. Growth and reproduction in captive tufted capuchins (*Cebus apella*). American Journal of Primatology 44, 197-203.

Fruhbeck, G., Jebb, S.A., Prentice, A.M., 1998. Leptin: physiology and pathophysiology. Clinical physiology (Oxford, England) 18, 399-419.

Galbreath, G., 1983. Karyotypic Evolution in *Aotus*. Am. J. Primatol. 4, 245-251.

Garber, P.A., Leigh, S.R., 1997. Ontogenetic variation in small-bodied New World primates: implications for patterns of reproduction and infant care. Folia Primatol (Basel) 68, 1-22.

Garcia-Mayor, R.V., Andrade, M.A., Rios, M., Lage, M., Dieguez, C., Casanueva, F.F., 1997. Serum leptin levels in normal children: relationship to age, gender, body mass index, pituitary-gonadal hormones, and pubertal stage. The Journal of clinical endocrinology and metabolism 82, 2849-2855.

Gebo, D.L., MacLatchy, L., Kityo, R., Deino, A., Kingston, J., Pilbeam, D., 1997. A hominoid genus from the early Miocene of Uganda. Science (New York, N.Y.) 276, 401-404.

Genain, C.P., Hauser, S.L., 2001. Experimental allergic encephalomyelitis in the New World monkey *Callithrix jacchus*. Immunol Rev 183, 159-172.

Gengozian, N., Batson, J.S., 1975. Single-born marmosets without hemopoietic chimerism: naturally occurring and induced. Journal of medical primatology 4, 252-261.

Gengozian, N., Batson, J.S., Eide, P., 1964. Hemaologic and cytogenetic evidence for chimerism in the marmoset, *Tamarinus nigricollis*. SAM-TDR-64-61. AMD TR Rep, 1-0.

Gingerich, P., 1980. Eocene Adapidae, paleobiogeography, and the origin of South American Platyrrhini. . In: Ciochon, R.L., Chiarelli, A.B. (Eds.), Evolutionary Biology of the New World Monkeys and Continental Drift. Plenum, New York, pp. 123-138.

Gingerich, P.D., Uhen, M.D., 1994. Time of origin of primates. Journal of human evolution 27, 443-445.

Goldani, A., Carvalho, G.S., Bicca-Marques, J.C., 2006. Distribution patterns of Neotropical primates (Platyrrhini) based on Parsimony Analysis of Endemicity. Brazilian journal of biology = Revista brasleira de biologia 66, 61-74.

Goldized, W.W., 1987. Tamarins and marmosets: communal care of offspring. In: Smutts BB. (Ed.), Primate Societies. University of Chicago Press, Chicago IL., pp. 34-43.

Goodman, M., 1961. The role of immunochemical differences in the phyletic development of human behavior. Human biology 33, 131-162.

Goodman, M., 1962. Evolution of the immunologic species specificity of human serum proteins. Human biology 34, 104-150.

Goodman, M., 1999. The genomic record of Humankind's evolutionary roots. Am J Hum Genet 64, 31-39.

Goodman, M., Grossman, L.I., Wildman, D.E., 2005. Moving primate genomics beyond the chimpanzee genome. Trends Genet 21, 511-517.

Goodman, M., Page, S., Meireles, C., Czelusniak, J., 1999. Primate Phylogeny and Classification Elucidated at the Molecular Level. In: Wasser, S. (Ed.), Evolutionary theory and processes: modern perspectives. Kluwer Academic Publishers, Dordrecht, p. 193*211.

Goodman, M., Porter, C.A., Czelusniak, J., Page, S.L., Schneider, H., Shoshani, J., Gunnell, G., Groves, C.P., 1998. Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9, 585-598.

Gradstein, F.M., Ogg, J.G., Smith, A.G., 2004. A geologic time scale 2004. Cambridge university press, Cambridge.

Green, P., 2007. 2x genomes--does depth matter? Genome research 17, 1547-1549.

Groves, C., 1986. Systematics of the great apes. In: Swindler, D., Erwin, J. (Eds.), Comparative primate biology: systematics, evolution, and anatomy. A.R. Liss, New York, pp. 187-217.

Groves, C., 2001. Primate Taxonomy. Smithsonian Institute Press, Washington DC.

Groves, C., 2005. Order Primates. In: Wilson, D., Reeder, D. (Eds.), Mammal Species ofthe World. John Hopkins University Press, Baltimore MD., pp. 111-184.

Haig, D., 1999. What is a marmoset? Am J Primatol 49, 285-296.

Halaas, J.L., Gajiwala, K.S., Maffei, M., Cohen, S.L., Chait, B.T., Rabinowitz, D., Lallone, R.L., Burley, S.K., Friedman, J.M., 1995. Weight-reducing effects of the

plasma protein encoded by the obese gene. Science (New York, N.Y.) 269, 543-546.

Hallstrom, B.M., Kullberg, M., Nilsson, M.A., Janke, A., 2007. Phylogenomic data analyses provide evidence that Xenarthra and Afrotheria are sister groups. Mol Biol Evol 24, 2059-2068.

Hamann, A., Matthaei, S., 1996. Regulation of energy balance by leptin. Experimental and clinical endocrinology & diabetes : official journal, German Society of Endocrinology [and] German Diabetes Association 104, 293-300.

Hamilton, W.D., 1964. The genetical evolution of social behaviour. I. J Theor Biol 7, 1-16.

Harada, M.L., Schneider, H., Schneider, M.P., Sampaio, I., Czelusniak, J., Goodman, M., 1995. DNA evidence on the phylogenetic systematics of New World monkeys: support for the sister-grouping of Cebus and Saimiri from two unlinked nuclear genes. Mol Phylogenet Evol 4, 331-349.

Harrison, T., 2002. Late Oligocene to middle Miocenecatarrhines from Afro-Arabia. In: Hartwig, W. (Ed.), Primate Fossile Record. Cambridge University Press, Cambridge MA., pp. 311-338.

Hartenberger, J.L., 1982. A review of the Eocene rodents of Pakistan. Contributions from the Museum of Paleontology, University of Michigan 26, 19-35.

Hartwig, W., Rosenberger, A.L., Norconk, M.A., Owl, M.Y., 2011. Relative brain size, gut size, and evolution in New World monkeys. Anatomical record (Hoboken, N.J. : 2007) 294, 2207-2221.

Hartwig, W.C., Meldrum, D.J., 2002. Miocene platyrrhines of the northern Neotropics. In: Hartwig, W.C. (Ed.), The Primate Fossil Record. Cambridge University Press, Cambridge.

Hasegawa, M., Kishino, H., Hayasaka, K., Horai, S., 1990. Mitochondrial DNA evolution in primates: transition rate has been extremely low in the lemur. Journal of molecular evolution 31, 113-121.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160-174.

Hayasaka, K., Gojobori, T., Horai, S., 1988. Molecular phylogeny and evolution of primate mitochondrial DNA. Mol. Biol. Evol. 5, 626-644.

Hendrickson, A., Djajadi, H.R., Nakamura, L., Possin, D.E., Sajuthi, D., 2000. Nocturnal tarsier retina has both short and long/medium-wavelength cones in an unusual topography. J Comp Neurol 424, 718-730.

Hershkovitz, P., 1977. Living New World Monkeys (Platyrrhini), Volume 1: With an Introduction to Primates. Chicago University Press, Chicago.

Hershkovitz, P., 1983. Two new species of night monkeys, genus *Aotus* (Cebidae, Platyrrhini): A preliminary report on Aotus taxonomy. Am. J. Primatol. 4, 209-243.

Hill, J.P., 1932. The Developmental History of the Primates. Philosophical Transactions of the Royal Society of London B 221, 45-178.

Hill, W., 1957. Primates: Comparative Anatomy and Taxonomy Vol 3: Pithecoidea, Platyrhini. Edinburgh University Press, Edinburgh.

Hirsch, A., Dias, L.G., Martins, L.d.O., Campos, R.F., Resende, N.A.T., Landau, E.C., In Prep. Database of Georeferenced Occurrence Localities of Neotropical Primates. Department of Zoology / UFMF, Belo Horizonte. www.icb.ufmg.br/zoo/primates/bdp_intro.htm.

Holmes, A.W., Devine, J.A., Nowakowski, E., Deinhardt, F., 1966. The epidemiology of a herpes virus infection of New World monkeys. J Immunol 96, 668-671.

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sanmartín, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghiemstra, H., Lundberg, J., Stadler, T., Särkinen, T., Antonelli, A., 2010. Amazonia Through Time: Andean Uplift, Climate Change, Landscape Evolution, and Biodiversity. Science (New York, N.Y.) 330, 927-931.

Horovitz, I., Meyer, A., 1995. Systematics of new world monkeys (Platyrrhini, primates) based on 16S mitochondrial DNA sequences: A comparative analysis of different weighting methods in cladistic analysis. Molecular Phylogenetics and Evolution 4, 448-456.

Houle, A., 1999. The origin of platyrrhines: An evaluation of the Antarctic scenario and the floating island model. American journal of physical anthropology 109, 541-559.

Huelsenbeck, J.P., 1995. Performance of Phylogenetic Methods in Simulation. Syst. Biol. 44, 17-48.

Hughes, S., Arneson, N., Done, S., Squire, J., 2005. The use of whole genome amplification in the study of human disease. Progress in biophysics and molecular biology 88, 173-189.

Inoue, J., Donoghue, P.C., Yang, Z., 2010. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. Syst Biol 59, 74-89.

IUCN, 2012. IUCN Red List of Threatened Species. Version 2012.1. www.iucnredlist.org.

Ivy, L.D., 1990. Systematics of Late Paleocene and Early Eocene Rodentia (Mammalia) from the Clarks Fork basin, Wyoming. Contributions from the Museum of Paleontology, University of Michigan 28, 21-70.

Izard, M.K., Wright, P.C., Simons, E.L., 1985. Gestation length in *Tarsius bancanus*. Am. J. Primatol. 9, 327-331.

Jacobs, L.L., Flynn, L.J., 2005. Of mice ... Again: The Siwalik Rodent Record, Murine Distribution, and Molecular Clocks. In: Lieberman, D.E., Smith, R.J., Kelley, J.

(Eds.), Interpreting the past: essays on Human, Primate, and Mammal evolution in honor of David Pilbeam. Brill Academic Publishers, Boston, pp. 63-80.

Jameson, N.M., Xu, K., Yi, S.V., Wildman, D.E., 2012. Development and annotation of shotgun sequence libraries from New World monkeys. Molecular ecology resources 12, 950-955.

Janis, C.M., 1998. Evolution of Terriary Mammals of North America. Cambridge University Press, Cambridge.

Jaquish, C.E., Tardif, S.D., Toal, R.L., Carson, R.L., 1996. Patterns of prenatal survival in the common marmoset (*Callithrix jacchus*). Journal of medical primatology 25, 57-63.

Jaworski, C.J., 1995. A reassessment of mammalian alpha A-crystallin sequences using DNA sequencing: implications for anthropoid affinities of tarsier. Journal of molecular evolution 41, 901-908.

Jones, K.E., Jon, B., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L., Safi, K., Sechrest, W., Boakes, E.H., Carbone, C., Connolly, C., Cutts, M.J., Foster, J.K., Grenyer, R., Habib, M., Plaster, C.A., Price, S.A., Rigby, E.A., Rist, J., Teacher, A., Bininda-Emonds, O.R.P., Gittleman, J.L., Mace, G.M., Purvis, A., 2009. PanTHERIA: A Species-Level Database of Life History, Ecology, and Geography of Extant and Recently Extinct Mammals. Ecology 90, 2648.

Jurke, M.H., Pryce, C.R., Dobeli, M., Martin, R.D., 1994. Nonincasive detection and monitoring of pregnancy and the postpartum period in Goeldi's monky (*Callimico*

goeldii) using urinary pregnanediol-3-alpha-glucuronide. American Journal of Primatology 34, 319-331.

Kappeler, P.M., Pereira, M.E., Schaik, C.P.v., 2003. Primate Life Histories and Socioecology. In: Kappeler, P.M., Pereira, M.E. (Eds.), Primate Life Histories and Socioecology. University of Chicago Press, Chicago, pp. 1-20.

Kass, R.E., Raftery, A.E., 1995. BAYES FACTORS. J. Am. Stat. Assoc. 90, 773-795.

Kay, R.F., 1990. The phyletic relationships of extant and fossil Pitherciinae (Platyrrhini, Anthropoidea). Journal of human evolution 19, 175-208.

Kay, R.F., 1994. Giant tamarin from the micoene of Colombia. American Journal of Physical Anthropology 95, 333-353.

Kay, R.F., Johnson, D., Meldrum, D.J., 1998. A new pitheciin primate from the middle Miocene of Argentina. Am. J. Primatol. 45, 317-336.

Kay, R.F., Perry, J., Malinzak, M., Allen, K.L., Kirk, E.C., Plavcan, J.M., Fleagle, J.G., 2012. Paleobiology of Santacrucian primates In: Sergio F. Vizcaino, Richard F. Kay and M. Susana Bargo (Eds.), Early Miocene Paleobiology in Patagonia: High-Latitude Paleocommunities of the Santa Cruz Formation. Cambridge University Press, Cambridge.

Kirkwood, J., 1983. Effects of diet on health, weight and litter-size in captive cotton-top tamarins *Saguinus oedipus oedipus*. Primates 24, 515-520.

Koop, B.F., Siemieniak, D., Slightom, J.L., Goodman, M., Dunbar, J., Wright, P.C., Simons, E.L., 1989. *Tarsius* delta- and beta-globin genes: conversions, evolution, and systematic implications. J Biol Chem 264, 68-79.

Landy, H.J., Weiner, S., Corson, S.L., Batzer, F.R., Bolognese, R.J., 1986. The "vanishing twin": ultrasonographic assessment of fetal disappearance in the first trimester. Am J Obstet Gynecol 155, 14-19.

Latrubesse, E.M., Cozzuol, M., da Silva-Caminha, S.A.F., Rigsby, C.A., Absy, M.L., Jaramillo, C., 2010. The Late Miocene paleogeography of the Amazon Basin and the evolution of the Amazon River system. Earth-Sci. Rev. 99, 99-124.

Lebatard, A.E., Bourles, D.L., Duringer, P., Jolivet, M., Braucher, R., Carcaillet, J., Schuster, M., Arnaud, N., Monie, P., Lihoreau, F., Likius, A., Mackaye, H.T., Vignaud, P., Brunet, M., 2008. Cosmogenic nuclide dating of Sahelanthropus tchadensis and *Australopithecus bahrelghazali*: Mio-Pliocene hominids from Chad. Proc Natl Acad Sci U S A 105, 3226-3231.

Leutenegger, W., 1979. Evolution of litter size in primates. American Naturalist 114, 525-531.

Li, W.H., Wu, C.I., Luo, C.C., 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2, 150-174.

Lovmar, L., Fredriksson, M., Liljedahl, U., Sigurdsson, S., Syvanen, A.C., 2003. Quantitative evaluation by minisequencing and microarrays reveals accurate

multiplexed SNP genotyping of whole genome amplified DNA. Nucleic acids research 31, e129.

Luckett, W.P., 1974. Reproductive Biology of the Primates. S. Karger.

Lynch Alfaro, J.W., Silva, J.D.E.S.E., Rylands, A.B., 2012. How Different Are Robust and Gracile Capuchin Monkeys? An Argument for the Use of *Sapajus* and *Cebus*. Am. J. Primatol. 74, 273-286.

MacFadden, B.J., 1990. Chronology of cenozoic primate localities in South America. Journal of human evolution 19, 151-156.

Macphee, R.D.E., Horovitz, I., 2002. Extinct Quaternary platyrrhines of the Greater Antilles and Brazil. In: Hartwig, W.C. (Ed.), The Primate Fossil Record. Cambridge University Press, Cambridge.

Maddison, W.P., 1997. Gene Trees in Species Trees. Syst. Biol. 46, 523-536.

Maddison, W.P., Donoghue, M.J., Maddison, D.R., 1984. Outgroup analysis and parsimony. Systematic Zoology 33, 83-103.

Mantzoros, C.S., Flier, J.S., Rogol, A.D., 1997. A longitudinal assessment of hormonal and physical alterations during normal puberty in boys. V. Rising leptin levels may signal the onset of puberty. The Journal of clinical endocrinology and metabolism 82, 1066-1070.

Marroig, G., Cheverud, J.M., 2001. A Comparison of phenotypic variation and covariation patterns and the role of phylogeny, ecology, and ontogeny during cranial evolution of New World monkeys. Evolution 55.

Martin, R.D., 1973. Comparitive anatomy and primate systematics. Symp. Zool. Soc. Lond. 33, 301-337.

Martin, R.D., 1990. Primate Origins and Evolution: a phylogenetic reconstruction. Princeton University Press, Princeton NJ.

Martin, R.D., 1992. Goeldi and the dwarfs: the evolutionary biology of the small New World monkeys. Journal of human evolution 22, 367-393.

Matsui, A., Rakotondraparany, F., Munechika, I., Hasegawa, M., Horai, S., 2009. Molecular phylogeny and evolution of prosimians based on complete sequences of mitochondrial DNAs. Gene 441, 53-66.

McNatty, K.P., Henderson, K.M., 1987. Gonadotrophins, fecundity genes and ovarian follicular function. Journal of steroid biochemistry 27, 365-373.

McNiff, B.E., Allard, M.W., 1998. A test of Archonta monophyly and the phylogenetic utility of the mitochondrial gene 12S rRNA. American journal of physical anthropology 107, 225-241.

Mein, P., Ginsburg, L., 1997. The mammals of the Lower Miocene deposits of Li Mae Long, Thailand: Systematics, biostratigraphy and paleoenvironments. Geodiversitas 19, 783-844.

Meireles, C., Czelusniak, J., Page, S.L., Wildman, D.E., Goodman, M., 2003. Phylogenetic Position of Tarsiers within the Order Primates: Evidence from γ-Globin DNA Sequences. In: Wright, P., Simons, E., Gursky, S. (Eds.), Tarsiers Past, Present and Future. Rutgers University Press, Piscataway NJ.

Menezes, A.N., Bonvicino, C.R., Seuanez, H.N., 2010. Identification, classification and evolution of owl monkeys (*Aotus, Illiger* 1811). BMC evolutionary biology 10, 248.

Merker, S., Groves, C.P., 2006. *Tarsius lariang*: a new primate species from western central Sulawesi. International Journal of Primatology 27, 465-485.

Miller, E.R., Gunnell, G.F., Martin, R.D., 2005. Deep time and the search for anthropoid origins. American journal of physical anthropology Suppl 41, 60-95.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, pp. 1-8.

Mittermeier, R.A., Ratsimbazafy, J., Rylands, A.B., Williamson, L., Oates, J.F., Mbora, D., Ganzhorn, J.U., Rodriguez-Luna, E., Palacios, E., Heymann, E.W., Kierulff, M.C.M., Long, Y.C., Supriatna, J., Roos, C., Walker, S., Aguiar, J.M., 2007. Primates in peril: the world's 25 most endangered primates, 2006-2008. Primate Conservation 22, 1-40.

Montgomery, G.W., Galloway, S.M., Davis, G.H., McNatty, K.P., 2001. Genes controlling ovulation rate in sheep. Reproduction (Cambridge, England) 121, 843-852.

Montgomery, G.W., McNatty, K.P., Davis, G.H., 1992. Physiology and molecular genetics of mutations that increase ovulation rate in sheep. Endocr Rev 13, 309-328.

Moore, H.D., Gems, S., Hearn, J.P., 1985. Early implantation stages in the marmoset monkey (*Callithrix jacchus*). Am J Anat 172, 265-278.

Morales-Jimenez, A., de la Torre, S., 2008. Aotus lemurinus. IUCN 2011, IUCN Red List of Threatened Species.

Moreira, M.A., 2002. SRY evolution in Cebidae (Platyrrhini: Primates). Journal of molecular evolution 55, 92-103.

Mounzih, K., Lu, R., Chehab, F.F., 1997. Leptin treatment rescues the sterility of genetically obese ob/ob males. Endocrinology 138, 1190-1193.

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.

Napier, J., Napier, P., 1985. The natural history of primates. MIT Press, Cambridge MA.

Nash, L.T., 1986. Dietary, behavioral, and morphological aspects of gummivory in primates. American journal of physical anthropology 29, 113-137.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, USA.

Newton, M.A., Raftery, A.E., Davison, A.C., Bacha, M., Celeux, G., Carlin, B.P., Clifford, P., Lu, C., Sherman, M., Tanner, M.A., Gelfand, A.E., Mallick, B.K., Gelman, A., Grieve, A.P., Kunsch, H.R., Leonard, T., Hsu, J.S.J., Liu, J.S., Rubin, D.B., Lo, A.Y., Louis, T.A., Neal, R.M., Owen, A.B., Tu, D.S., Gilks, W.R., Roberts, G., Sweeting, T., Bates, D., Ritter, G., Worton, B.J., Barnard, G.A., Gibbens, R., Silverman, B., 1994. Approximate Bayesian-inference with the weithted likelihoos bootstrap. J. R. Stat. Soc. Ser. B-Methodol. 56, 3-48.

Niblack, G.D., Kateley, J.R., Gengozian, N., 1977. T-and B-lymphocyte chimerism in the marmoset. Immunology 32, 257-263.

Niemitz, C., 1983. New Results on the locomothion of *Tarsius bancanus*, horsfiels, 1821. Annales Des Sciences Naturelles-Zoologie Et Biologie Animale 5, 89-100.

Nievergelt, C.M., Mundy, N.I., Woodruff, D.S., 1998. Microsatellite primers for genotyping common marmosets (*Callithrix jacchus*) and other callitrichids. Molecular Ecology 7, 1432-1434.

Nieves, M., Ascunce, M.S., Rahn, M.I., Mudry, M.D., 2005. Phylogenetic relationships among some *Ateles* species: the use of chromosomic and molecular characters. Primates 46, 155-164.

Nikolaev, S.I., Montoya-Burgos, J.I., Popadin, K., Parand, L., Margulies, E.H., Antonarakis, S.E., 2007. Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. Proc Natl Acad Sci U S A 104, 20443-20448.

Nishihara, H., Okada, N., Hasegawa, M., 2007. Rooting the eutherian tree: the power and pitfalls of phylogenomics. Genome Biol 8, R199.

Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of molecular biology 302, 205-217.

Nylander, J.A.A., 2004. MrModeltest v2. Evolutionary Biology Center, Uppsala University.

Ohta, T., 1998. Evolution by nearly-neutral mutations. Genetica 102-103, 83-90.

Ollivier, F.J., Samuelson, D.A., Brooks, D.E., Lewis, P.A., Kallberg, M.E., Komaromy, A.M., 2004. Comparative morphology of the tapetum lucidum (among selected species). Vet Ophthalmol 7, 11-22.

Onichtchouk, D., Chen, Y.G., Dosch, R., Gawantka, V., Delius, H., Massague, J., Niehrs, C., 1999. Silencing of TGF-beta signalling by the pseudoreceptor BAMBI. Nature 401, 480-485.

Opazo, J.C., Wildman, D.E., Prychitko, T., Johnson, R.M., Goodman, M., 2006. Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates). Mol Phylogenet Evol 40, 274-280.

Organ, J.M., Teaford, M.F., Taylor, A.B., 2009. Functional correlates of fiber architecture of the lateral caudal musculature in prehensile and nonprehensile

tails of the platyrrhini (primates) and procyonidae (carnivora). Anatomical record (Hoboken, N.J. : 2007) 292, 827-841.

Osterholz, M., Walter, L., Roos, C., 2009. Retropositional events consolidate the branching order among New World monkey genera. Mol Phylogenet Evol 50, 507-513.

Padula, A.M., 2005. The freemartin syndrome: an update. Anim Reprod Sci 87, 93-109.

Page, S.L., Goodman, M., 2001. Catarrhine phylogeny: noncoding DNA evidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade. Mol Phylogenet Evol 18, 14-25.

Pancer, Z., Gershon, H., Rinkevich, B., 1995. Cloning of a urochordate cDNA featuring mammalian short consensus repeats (SCR) of complement-control protein superfamily. Comp Biochem Physiol B Biochem Mol Biol 111, 625-632.

Papper, Z., Jameson, N.M., Romero, R., Weckle, A.L., Mittal, P., Benirschke, K., Santolaya-Forgas, J., Uddin, M., Haig, D., Goodman, M., Wildman, D.E., 2009. Ancient origin of placental expression in the growth hormone genes of anthropoid primates. Proc Natl Acad Sci U S A 106, 17083-17088.

Pastorini, J., Forstner, M.R., Martin, R.D., Melnick, D.J., 1998. A reexamination of the phylogenetic position of *Callimico* (primates) incorporating new mitochondrial DNA sequence data. Journal of molecular evolution 47, 32-41.

Pelleymounter, M.A., Cullen, M.J., Baker, M.B., Hecht, R., Winters, D., Boone, T., Collins, F., 1995. Effects of the obese gene product on body weight regulation in ob/ob mice. Science (New York, N.Y.) 269, 540-543.

Peng, Z., Elango, N., Wildman, D.E., Yi, S.V., 2009. Primate phylogenomics: developing numerous nuclear non-coding, non-repetitive markers for ecological and phylogenetic applications and analysis of evolutionary rate variation. BMC Genomics 10, 247.

Perelman, P., Johnson, W.E., Roos, C., Seuanez, H.N., Horvath, J.E., Moreira, M.A., Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., Schneider, M.P., Silva, A., O'Brien, S.J., Pecon-Slattery, J., 2011. A molecular phylogeny of living primates. PLoS genetics 7, e1001342.

Perez, S.I., Klaczko, J., dos Reis, S.F., 2012. Species tree estimation for a deep phylogenetic divergence in the New World monkeys (Primates: Platyrrhini). Molecular Phylogenetics and Evolution 65, 621-630.

Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T., Manuel, M., Worheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS biology 9, e1000602.

Plunkett, J., Doniger, S., Orabona, G., Morgan, T., Haataja, R., Hallman, M., Puttonen, H., Menon, R., Kuczynski, E., Norwitz, E., Snegovskikh, V., Palotie, A., Peltonen, L., Fellman, V., DeFranco, E.A., Chaudhari, B.P., McGregor, T.L., McElroy, J.J., Oetjens, M.T., Teramo, K., Borecki, I., Fay, J., Muglia, L., 2011. An evolutionary

genomic approach to identify genes involved in human birth timing. PLoS genetics 7, e1001365.

Pocock, R.I., 1918. On the external characters of the *Lemurs* and of *Tarsius*. Proc. Zool. Soc. Lond. 1918, 19-53.

Pons, D., De Franceschi, D., 2007. Neogene woods from western Peruvian Amazon and palaeoenvironmental interpretation. Bulletin of Geosciences 82, 343-354.

Pook, A., 1977. Comparison between the reproduction and parental behacious of the Goeldi's monkey (*Callimico goeldii*) and of the true marmosets (Callitrichidae). In: Rothe, H. (Ed.), Proceedings of the marmoset workshop. W. Germany, Gottingen.

Porter, C.A., Czelusniak, J., Schneider, H., Schneider, M.P., Sampaio, I., Goodman, M., 1997a. Sequences of the primate epsilon-globin gene: implications for systematics of the marmosets and other New World primates. Gene 205, 59-71.

Porter, C.A., Czelusniak, J., Schneider, H., Schneider, M.P., Sampaio, I., Goodman, M., 1999. Sequences from the 5' flanking region of the epsilon-globin gene support the relationship of *Callicebus* with the pitheciins. Am J Primatol 48, 69-75.

Porter, C.A., Page, S.L., Czelusniak, J., Schneider, H., Schneider, M.P.C., Sampaio, I., Goodman, M., 1997b. Phylogeny and Evolution of Selected Primates as Determined by Sequences of the ε-Globin Locus and 5′ Flanking Regions. International Journal of Primatology 18, 261-295.

Porter, C.A., Sampaio, I., Schneider, H., Schneider, M.P., Czelusniak, J., Goodman, M., 1995. Evidence on primate phylogeny from epsilon-globin gene sequences and flanking regions. Journal of molecular evolution 40, 30-55.

Poudyal, M., Rosa, S., Powell, A.E., Moreno, M., Dellaporta, S.L., Buss, L.W., Lakkis, F.G., 2007. Embryonic chimerism does not induce tolerance in an invertebrate model organism. Proc Natl Acad Sci U S A 104, 4559-4564.

Poulsen, C.J., Ehlers, T.A., Insel, N., 2010. Onset of convective rainfall during gradual late Miocene rise of the central Andes. Science (New York, N.Y.) 328, 490-493.

Poux, C., Chevret, P., Huchon, D., de Jong, W.W., Douzery, E.J., 2006. Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. Syst Biol 55, 228-244.

Prasad, A.B., Allard, M.W., Green, E.D., 2008. Confirming the phylogeny of mammals by use of large comparative sequence data sets. Mol Biol Evol 25, 1795-1808.

Pree, I., Pilat, N., Wekerle, T., 2007. Recent progress in tolerance induction through mixed chimerism. Int Arch Allergy Immunol 144, 254-266.

Promislow, D.E.L., Harvey, P.H., 1990. Living fast and dying young: A comparative analysis of life-history variation among mammals. Journal of Zoology 220, 417-437.

Rannala, B., Yang, Z., 2007. Inferring speciation times under an episodic molecular clock. Syst Biol 56, 453-466.

Ranwez, V., Harispe, S., Delsuc, F., Douzery, E.J.P., 2011. MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. PloS one 6, e22594.

Rasmussen, D.T., Conroy, G.C., Simons, E.L., 1998. Tarsier-like locomotor specializations in the oligocene primate afrotarsius. Proc Natl Acad Sci U S A 95, 14848-14850.

Raveendran, M., Tardif, S., Ross, C.N., Austad, S.N., Harris, R.A., Milosavljevic, A., Rogers, J., 2008. Polymorphic microsatellite loci for the common marmoset (*Callithrix jacchus*) designed using a cost- and time-efficient method. Am. J. Primatol. 70, 906-910.

Ray, D.A., Xing, J., Hedges, D.J., Hall, M.A., Laborde, M.E., Anders, B.A., White, B.R., Stoilova, N., Fowlkes, J.D., Landry, K.E., Chemnick, L.G., Ryder, O.A., Batzer, M.A., 2005. Alu insertion loci and platyrrhine primate phylogeny. Mol Phylogenet Evol 35, 117-126.

Roberts, M., 1994. Growth, development, and paternal care in the western tarsier (*Tarsius bancanus*) in captivity: evidence for a "slow" life-history and nonmonogamous mating system. Inernational Journal of Primatology 15, 1-28.

Robson, S.K., Rouse, G.W., Pettigrew, J.D., 1997. Sperm ultrastructure of *Tarsius bancanus* (Tarsiidae, Primates): Implications for primate phylogeny and the use of sperm in systematics. Acta Zool. 78, 269-278.

Roca, A.L., Bar-Gal, G.K., Eizirik, E., Helgen, K.M., Maria, R., Springer, M.S., O'Brien, S.J., Murphy, W.J., 2004. Mesozoic origin for West Indian insectivores. Nature 429, 649-651.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics (Oxford, England) 19, 1572-1574.

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. Proc Biol Sci 275, 1203-1208.

Rosenberger, A., 2002. Platyrrhine paleontology and systematics: the paradigm shifts. In: WC, H. (Ed.), The Primate Fossil Record. Cambridge University Press, U.K., pp. 151-159.

Rosenberger, A., Szalay, F., 1980. On the tarsiiform origins of Anthropoidea. In: Ciochon, R., Chiarelli, A. (Eds.), Evolutionary biology of the New World monkeys and continental drift. Plenum Press, New York, pp. 139-157.

Rosenberger, A.L., 1981. Systematics: the higher taxa. In: Coimbra-Filho A.F, M.R.A. (Ed.), Ecology and Behaviour of Neotropical Primates. Academia Brasileira de Ciencias, Rio de Janeiro, pp. 9-27.

Rosenberger, A.L., 1984. Fossil New World monkeys dispute the molecular clock. Journal of human evolution 13, 737-742.

Rosenberger, A.L., 2012. New World Monkey Nightmares: Science, Art, Use, and Abuse (?) in Platyrrhine Taxonomic Nomenclature. Am J Primatol.

Rosenberger, A.L., Coimbrafilho, A.F., 1984. Morphology, taxonomic statu and affinities od the lion tamarins, *Leontopithecus* (callitrichinae, Cebidae). Folia Primatologica 42, 149-179.

Rosenbloom, L.A., 1968. Some aspects of female reproductive physiology in the squirrel monkey. In: Rosenbloom, L.A., Cooper, R.W. (Eds.), The Squirrel monkey. Academic Press, New York, pp. 147-169.

Ross, A.C., Porter, L.M., Power, M.L., Sodaro, V., 2010. Maternal care and infant development in *Callimico goeldii* and *Callithrix jacchus*. Primates 51, 315-325.

Ross, C., 1991. Life history patterns of new world monkeys. International Journal of Primatology 12, 481-502.

Ross, C.N., French, J.A., Orti, G., 2007. Germ-line chimerism and paternal care in marmosets (*Callithrix kuhlii*). Proc Natl Acad Sci U S A 104, 6278-6282.

Rutherford, J.N., Tardif, S.D., 2008. Placental efficiency and intrauterine resource allocation strategies in the common marmoset pregnancy. American journal of physical anthropology 137, 60-68.

Rylands, A.B., Schneider, H., Langguth, A., Mittermeier, R.A., Groves, C.P., Rodriguez-Luna, E., 2000. An assessment of the diversity of New World primates. Neptrop. Primates 8, 61-93.

Schmitz, J., Ohme, M., Zischler, H., 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of Tarsius bancanus to other primates. Genetics 157, 777-784.

Schneider, H., Bernardi, J.A.R., Da Cunha, D.B., Tagliaro, C.H., Vallinoto, M., Ferrari, S.F., Sampaio, I., 2012. A molecular analysis of the evolutionary relationships in the Callitrichinae, with emphasis on the position of the dwarf marmoset. Zoologica Scripta 41, 1-10.

Schneider, H., Canavez, F.C., Sampaio, I., Moreira, M.A.M., Tagliaro, C.H., Seuanez, H.N., 2001. Can molecular data place each neotropical monkey in its own branch? Chromosoma 109, 515-523.

Schneider, H., Sampaio, I., Harada, M.L., Barroso, C.M.L., Schneider, M.P.C., Czelusniak, J., Goodman, M., 1996. Molecular phylogeny of the new world monkeys (Platyrrhini, primates) based on two unlinked nuclear genes: *IRBP* intron 1 and epsilon-globin sequences. American journal of physical anthropology 100, 153-179.

Schneider, H., Schneider, M.P., Sampaio, I., Harada, M.L., Stanhope, M., Czelusniak, J., Goodman, M., Schneider, H., Schneider, M.P., Sampaio, M.I., Montoya, E., Tapia, J., Encarnacion, F., Anselmo, N.P., Salzano, F.M., 1993. Molecular phylogeny of the New World monkeys (Platyrrhini, primates) Divergence between biochemical and cytogenetic differences in three species of the *Callicebus moloch* group. Mol Phylogenet Evol 2, 225-242.

Schradin, C., Anzenberger, G., 2001. Infant carrying in family groups of goeldi's monkeys (*Callimico goeldii*). Am J Primatol 53, 57-67.

Schwartz, J.H., 2003. How Close Are the Similarities between *Tarsius* and Other Primates. In: Wright, P.C., Simons, E.L., Gursky, S. (Eds.), Tarsiers: Past, Present, and Future. Rutgers University Press, New Brunswick, NJ.

Seiffert, E.R., 2006. Revised age estimates for the later Paleogene mammal faunas of Egypt and Oman. Proc. Natl. Acad. Sci. USA 103, 5000-5005.

Seo, T.K., Kishino, H., Thorne, J.L., 2004. Estimating absolute rates of synonymous and nonsynonymous nucleotide substitution in order to characterize natural selection and date species divergences. Mol Biol Evol 21, 1201-1213.

Seuanez, H.N., Forman, L., Alves, G., 1988. Comparative chromosome morphology in three Callitrichid genera: *Cebuella*, *Callithrix*, and *Leontopithecus*. J Hered 79, 418-424.

Shimasaki, S., Moore, R.K., Otsuka, F., Erickson, G.F., 2004. The Bone Morphogenetic Protein System In Mammalian Reproduction. Endocrine Reviews 25, 72-101.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114-1116.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., Weinstock, G.M., Wilson, R.K., Gibbs, R.A., Kent, W.J., Miller, W., Haussler, D., 2005. Evolutionarily

conserved elements in vertebrate, insect, worm, and yeast genomes. Genome research 15, 1034-1050.

Silcox, M.T., 2008. The biogeographic origins of Primates and Euprimates: east, west, north or south of Eden? In: Sargis, E.J., Dagosto, M. (Eds.), Mammalian evolutionary morphology: a tribute to Fredrick S. Szalay. Springer, Dordrecht, The Netherlands.

Simons, E., Brown, T., 1985. *Afrotarsius chatrathi*, first tarsiiform primate (? Tarsiidae) from Africa. Nature 313, 475-477.

Simons, E.L., 1997. Discovery of the smallest Fayum Egyptian primates (Anchomomyini, Adapidae). Proceedings of the National Academy of Sciences 94, 180-184.

Simpson, G.G., 1945. The principles of classification and a classification of the mammals. Bull. Am. Mus. Nat. Hist. 85, 1-350.

Simpson, G.G., 1978. Early mammals in South America: Fact, controversy and mystery. Proc. Am. Philos. Soc. 122, 318-328.

Smit, A., Hunley, R, Green, P, 1996-2004. Repeatmasket Open-3.0.

Snowdon, 1993. A vocal taxonomy of the callitrichids. In: Rylands, A.B. (Ed.), Marmosets and Tamarins. Systematics Behaviour and Ecology. Oxford University Press, New York, pp. 78-94.

Springer, M.S., Meredith, R.W., Gatesy, J., Emerling, C.A., Park, J., Rabosky, D.L., Stadler, T., Steiner, C., Ryder, O.A., Janečka, J.E., Fisher, C.A., Murphy, W.J., 2012. Macroevolutionary Dynamics and Historical Biogeography of Primate Diversification Inferred from a Species Supermatrix. PloS one 7, e49521.

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

Stanyon, R., Bonvicino, C.R., Svartman, M., Seuanez, H.N., 2003. Chromosome painting in *Callicebus lugens*, the species with the lowest diploid number (2n=16) known in primates. Chromosoma 112, 201-206.

Starzl, T.E., Demetris, A.J., Trucco, M., Murase, N., Ricordi, C., Ildstad, S., Ramos, H., Todo, S., Tzakis, A., Fung, J.J., et al., 1993. Cell migration and chimerism after whole-organ transplantation: the basis of graft acceptance. Hepatology 17, 1127-1152.

Steiper, M.E., Young, N.M., 2006. Primate molecular divergence dates. Mol Phylogenet Evol 41, 384-394.

Steiper, M.E., Young, N.M., 2008. Timing Primate Evolution: Lessons from the Discordance Between Molecular and Paleontological Estimates. Evol. Anthro. 17, 179-188.

Stewart, C.B., Disotell, T.R., 1998. Primate evolution - in and out of Africa. Current biology : CB 8, R582-588.

Suchard, M.A., Weiss, R.E., Sinsheimer, J.S., 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol. Biol. Evol. 18, 1001-1013.

Sussman, R.W., Kinzey, W.G., 1984. The ecological role of the callitrichidae: a review. American journal of physical anthropology 64, 419-449.

Swofford, D., 2002. PAUP* Phylogenetic Analysis using Parsimony (and other methods). Sinauer Associates.

Tabuce, R., Marivaux, L., Adaci, M., Bensalah, M., Hartenberger, J.L., Mahboubi, M., Mebrouk, F., Tafforeau, P., Jaeger, J.J., 2007. Early Tertiary mammals from North Africa reinforce the molecular Afrotheria clade. Proc Biol Sci 274, 1159-1166.

Takai, M., 1994. New specimens of *Neosaimiri fieldsi* from La Venta, Colombia: a middle Miocene ancestor of the living squirrel monkeys. Journal of human evolution 27, 329-360.

Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56, 564-577.

Tardif, S., Power, M., Layne, D., Smucny, D., Ziegler, T., 2004. Energy restriction initiated at different gestational ages has varying effects on maternal weight gain and pregnancy outcome in common marmoset monkeys (*Callithrix jacchus*). The British journal of nutrition 92, 841-849.

Tardif, S.D., Harrison, M., Simek, M., 1993a. Communal infant care in marmosets and tamarins: relation to energetics, ecology and social organization. In: Rylands, A.B. (Ed.), Marmosets and tamarins: systematics behaviour and ecology. Oxford University Press, Oxford, pp. 220-234.

Tardif, S.D., Jaquish, C.E., 1997. Number of ovulations in the marmoset monkey (*Callithrix jacchus*): relation to body weight, age and repeatability. Am J Primatol 42, 323-329.

Tardif, S.D., Lacker, H.M., Feuer, M., 1993b. Follicular development and ovulation in the marmoset monkey as determined by repeated laparoscopic examination. Biology of reproduction 48, 1113-1119.

Tardif, S.D., Ross, C.N., 2009. Integration of proximate and evolutionary explanation of reproductive strategy: the case of callitrichid primates and implications for human biology. American journal of human biology : the official journal of the Human Biology Council 21, 731-738.

Tardif, S.D., Smucny, D.A., Abbott, D.H., Mansfield, K., Schultz-Darken, N., Yamamoto, M.E., 2003. Reproduction in captive common marmosets (*Callithrix jacchus*). Comparative Medicine 53, 364-368.

Terhune, C.E., 2011. Dietary correlates of temporomandibular joint morphology in New World primates. Journal of human evolution 61, 583-596.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence

weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research 22, 4673-4680.

Torres, O.M., Enciso, S., Ruiz, F., Silva, E., Yunis, I., 1998. Chromosome diversity of the genus *Aotus* from Colombia. Am J Primatol 44, 255-275.

Valenzuela, N., 1994. Early behavioral development of three wild infant *cebus apella* in Columbia. In: Roeder, J., Thierry, B., ANderson, J., Herrenschmidt, N. (Eds.), Current Primitology, Social development learning and behavior. Univ Louis Pasteur, Strasbourg.

Veiga, L., Bowler, M., Silva, J.J., Queiroz, H., Boubli, J.-P., Rylands, A.B., 2008. *Cacajao calvus*. In: 2011, I. (Ed.), IUCN Red List of Threatened Species. Version 2011.2.

Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., Birney, E., 2009. Ensembl Compara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome research 19, 327-335.

Visel, A., Prabhakar, S., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Afzal, V., Rubin, E.M., Pennacchio, L.A., 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nature genetics 40, 158-160.

von Dornum, M., Ruvolo, M., 1999. Phylogenetic relationships of the New World monkeys (Primates, platyrrhini) based on nuclear G6PD DNA sequences. Mol Phylogenet Evol 11, 459-476.

von Roosmalen, M.G.M., von Rossmalen, T., 2003. The description of a new marmoset genus, *Callibella* (Callitrichinae, Primates,) including it's molecular phylogenetic status. Neotropical Primates 11, 1-10.

Wang, Q.F., Prabhakar, S., Chanan, S., Cheng, J.F., Rubin, E.M., Boffelli, D., 2007. Detection of weakly conserved ancestral mammalian regulatory sequences by primate comparisons. Genome Biol 8, R1.

Wang, X.P., Yu, L., Roos, C., Ting, N., Chen, C.P., Wang, J., Zhang, Y.P., 2012. Phylogenetic relationships among the colobine monkeys revisited: new insights from analyses of complete mt genomes and 44 nuclear non-coding markers. PloS one 7, e36274.

Webb, S.D., 1991. Ecogeography and the great american interchange. Paleobiology 17, 266-280.

Werman, S.D., 2005. Hypotheses on the Historical Biogeography of Bothropid Pitvipers and Related Genera of the Neotropics. In: Donnely, M., Crother, B., Guyer, C., Wake, M., White, M. (Eds.), Ecology and Evolution in the Tropics: A Herpetological Perspective University of Chicago Press, Chicago IL.

Wesley-Hunt, G.D., Flynn, J.J., 2005. Phylogeny of the Carnivora: basal relationships among the Carnivoramorphans, and assessment of the position of 'Miacoidea' relative to Carnivora. Journal of Systematic Palaeontology 3, 1-28.

Wesselingh, F.P., Salo, J.A., 2006. A Miocene perspective on the evolution of the Amazonian biota. Scripta Geol 133, 439-458.

Wiens, J.J., Kuczynski, C.A., Smith, S.A., Mulcahy, D.G., Sites, J.W., Townsend, T.M., Reeder, T.W., 2008. Branch Lengths, Support, and Congruence: Testing the Phylogenomic Approach with 20 Nuclear Loci in Snakes. Syst. Biol. 57, 420-431.

Wildman, D.E., Bergman, T.J., al-Aghbari, A., Sterner, K.N., Newman, T.K., Phillips-Conroy, J.E., Jolly, C.J., Disotell, T.R., 2004. Mitochondrial evidence for the origin of hamadryas baboons. Mol Phylogenet Evol 32, 287-296.

Wildman, D.E., Chen, C., Erez, O., Grossman, L.I., Goodman, M., Romero, R., 2006. Evolution of the mammalian placenta revealed by phylogenetic analysis. Proc Natl Acad Sci U S A 103, 3203-3208.

Wildman, D.E., Goodman, M., 2004. Humankind's place in a phylogenetic classification of living primates. In: Wasser, S. (Ed.), Evolutionary Theory and Processes: Modern Horizons, Papers in Honor to Eviatar Nevo. Kluwer Academic Publishers, Netherlands, pp. 293-311.

Wildman, D.E., Jameson, N.M., Opazo, J.C., Yi, S.V., 2009. A fully resolved genus level phylogeny of neotropical primates (Platyrrhini). Mol Phylogenet Evol.

Wildman, D.E., Uddin, M., Opazo, J.C., Liu, G., Lefort, V., Guindon, S., Gascuel, O., Grossman, L.I., Romero, R., Goodman, M., 2007. Genomics, biogeography, and the diversification of placental mammals. Proc Natl Acad Sci U S A 104, 14395-14400.

Wilkinson, R.D., Steiper, M.E., Soligo, C., Martin, R.D., Yang, Z., Tavaré, S., 2011. Dating Primate Divergences through an Integrated Analysis of Palaeontological and Molecular Data. Syst Biol.

Williams, B.A., Kay, R.F., Kirk, E.C., 2010. New perspectives on anthropoid origins. Proceedings of the National Academy of Sciences.

Windle, C.P., Baker, H.F., Ridley, R.M., Oerke, A.K., Martin, R.D., 1999. Unrearable litters and prenatal reduction of litter size in the common marmoset (*Callithrix jacchus*). Journal of medical primatology 28, 73-83.

Wislocki, G., 1939. Observations on twinning in marmosets. Am. J. Anat. 64, 445-483.

Wislocki, G.B., 1929. On the placentation of primates, with a consideration of the phylogeny of the placenta. Contributions to Embryology 20, 53-U15.

Witte, S.M., Rogers, J., 1999. Microsatellite polymorphisms in Bolivian squirrel monkeys (*Saimiri boliviensis*). Am J Primatol 47, 75-84.

Wright, P.C., 1981. The night mokey, genus aotus. In: Coimbra-Filho, A.F., Mittermeier, R.A. (Eds.), Ecology and behavior of neotropical primates. Academic Brasileira de Ciencias, Rio de Janeiro, pp. 211-240.

Wright, P.C., Simons, E.L., Gursky, S., 2003. Tarsiers. Past, Present, and Future. Rutgers University Press, Piscataway NJ.

Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39, 306-314.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24, 1586-1591.

Yang, Z., Rannala, B., 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol 23, 212-226.

Yang, Z., Yoder, A.D., 1999. Estimation of the transition/transversion rate bias and species sampling. Journal of molecular evolution 48, 274-283.

Yoder, A.D., 2003. The phylogenetic position of the genus *Tarsius*: Whose side are you on? In: Wright, P.C., SImons, E.L., Gursky, S. (Eds.), Tarsiers Past, Present and Future. Rutgers university press, Piscataway NJ.

Yu, W.H., Kimura, M., Walczewska, A., Karanth, S., McCann, S.M., 1997. Role of leptin in hypothalamic-pituitary function. Proc Natl Acad Sci U S A 94, 1023-1028.

Yu, Y., Harris, A.J., He, X.-J., 2010. S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. Molecular Phylogenetics and Evolution 56, 848-850.

Yu, Y., Harris, A.J., He, X.-J., 2011. A Novel Bayesian Method for Reconstructing Geographic Ranges and Ancestral States on Phylogenies. Submitted.

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., Friedman, J.M., 1994. Positional cloning of the mouse obese gene and its human homologue. Nature 372, 425-432.

Zhao, G.Q., Hogan, B.L., 1996. Evidence that mouse *Bmp8a* (Op2) and *Bmp8b* are duplicated genes that play a role in spermatogenesis and placental development. Mechanisms of development 57, 159-168.

Zietkiewicz, E., Richer, C., Labuda, D., 1999. Phylogenetic affinities of tarsier in the context of primate Alu repeats. Mol Phylogenet Evol 11, 77-83.

ABSTRACT

PLATYRRHINE PHYLOGENTICS WITH A FOCUS ON CALLITRICHINE LIFE HISTORY ADAPTATIONS

by

NATALIE MAE JAMESON

December 2013

Advisor: Dr. Derek E. Wildman

Major: Molecular Genetics and Biology

Degree: Doctor of Philosophy

The life history of a species is highly impacted by their reproductive strategy. In my dissertation I address the changing reproductive strategies in callitrichine New World monkeys and their genetic underpinnings using a phylogenetic approach. The necessity for a resolved phylogeny is universal to any comparative genomic study. Here we have constructed a reliable phylogenetic framework from which reproductive strategy could be studied in callitrichines. First, to determine the most recent common ancestor of Anthropoid primates we took a phylogenomic approach, using the publicly available whole genome sequences of 17 mammal species. With high confidence, we determined here that Tarsier is the most recent common ancestor to Anthropoid primates. Secondly, we construct a reliable phylogenetic framework for New World monkeys. To do this, genomic sequence databases are developed and parsed for non-genic markers. The resulting phylogeny is based on 40+kb of non-genic genomic data and contains 40 species. Finally the reproductive strategy of callitrichines was investigated. The timing and mechanism of litter size reduction in Goeldi's monkey was accessed though detection of chimerism and adaptive evolution of genes involved in

reproduction. We determined based on these analysis that the reduction in litter size is likely pre-ovulation and due to a reversion to mono-ovulation in the species.

AUTOBIOGRAPHICAL STATEMENT

NATALIE MAE JAMESON

EDUCATION

PhD. Wayne State University School Of Medicine, Center for Molecular Medicine & Genetics, 2007–13
BA. Oakland University Department of Biological Sciences, 2006

FELLOWSHIPS AND AWARDS

2012-13 Provost Fellowship for Computational Biology, Wayne State University
2009 Bursery Award, Wellcome Trust Advanced Course
2009 Acceptance to Advanced Course in Molecular Evolution, Wellcome Trust, Hinxton, England
2004 Merck/AAAS Research Fellowship, Oakland University

GRANT SUPPORT

2011 Dissertation Improvement Grant, The National Science Foundation (BCS 1061370)
"Doctoral Dissertation Improvement: Development of genomic methods to clarify the evolution of life history trade-offs", $19,827

TEACHING

2009 Teaching assistant, Molecular Phylogenetics Course, Pontificia Universidad Javeriana, Bogota

COMMITTEES

2013 CMMG graduate student recruitment, student committee member
2010 Graduate Stusent Research Day, Chair of Funding
2010 Undergraduate Summer Research Program, Co-Chair
2009 Graduate Student Research Day, Funding committee member
2009 Undergraduate Summer Research Program, Co-Chair

SELECTED PUBLICATIONS

Jameson NM, Yi SV, Xu K, Wildman DE. **The tempo and mode of New World monkey evolution and biogeography in the context of phylogenomic analysis**. Molecular Phylogenetics and Evolution, In Review. PMID: TBD

Jameson NM, Xu K, Yi SV, Wildman DE. **Development and annotation of shotgun sequence libraries from New World monkeys** (2012). Molecular Ecology Resources. PMID: 22715851

Sterner KN, Jameson NM, Wildman DE. **Placental development, evolution, and epigenetics of primate pregnancies** (2012). In "Building Babies: Primate Development in Proximate and Ultimate Perspective" Eds. Julienne Rutherford, Katherine Clancy, and Katie Hinde, Springer.

Jameson NM, Hou ZC, Sterner KN, Weckle A, Goodman M, Steiper ME, Wildman DE. **Genomic data reject the hypothesis of a prosimian primate clade** (2011). Journal of Human Evolution. PMID: 21620437

Wildman DE, Jameson NM, Opazo JC, Yi SV, **A fully resolved genus level phylogeny of neotropical primates (Platyrrhini)** (2009). Molecular Phylogenetics and Evolution. PMID: 19632342

Papper Z, Jameson NM, Romero R, Weckle AL, Mittal P, Benirschke K, Santolaya-Forgas J, Uddin M, Haig D, Goodman M, Wildman DE. **Ancient origin of placental expression in the growth hormone genes of anthropoid primates.** PNAS. PMID: 19805162