MOLECULAR EVOLUTION OF VASOPRESSIN AND OXYTOCIN RECEPTOR GENES IN OWL MONKEYS (*AOTUS AZARAI*) OF NORTHERN ARGENTINA

Paul L. Babb

A DISSERTATION

in

Anthropology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2012

Supervisor of Dissertation

*Signature*_____

Theodore G. Schurr, Ph.D.
Associate Professor, Anthropology

Co-Supervisor

*Signature*_____

Eduardo Fernandez-Duque, Ph.D.
Assistant Professor, Anthropology

Graduate Group Chairperson

*Signature*_____

Deborah A. Thomas, Ph.D.
Professor of Anthropology

Dissertation Committee

Theodore G. Schurr, Ph.D., Associate Professor of Anthropology
Eduardo Fernandez-Duque, Ph.D., Assistant Professor of Anthropology
R. Arlen Price, Ph.D., Professor of Psychiatry

MOLECULAR EVOLUTION OF VASOPRESSIN AND OXYTOCIN RECEPTOR GENES IN OWL MONKEYS (*AOTUS AZARAI*) OF NORTHERN ARGENTINA

DEDICATION

For Effie and Alex

ACKNOWLEDGEMENTS

The author would like to express his gratitude to the many students, volunteers, and researchers who assisted with the capturing and sampling of owl monkeys that led to the collection of samples used in this genetic study. Special thanks are owed to Marcelo Rotundo for his assistance in the field during the past decade, as well as to Mr. John Adams, Mr. F. Middleton, and Ing. C. Cimino for their continuing support of the field site at Estancia Guaycolec in Formosa, Argentina. In addition, the author offers his many thanks to the University of Pennsylvania's Department of Anthropology and Graduate Group, the Graduate and Professional Student Assembly (GAPSA) at the University of Pennsylvania, and to the Leakey Foundation for project support over the last five years. The author is further indebted to his colleagues, Dr. Ömer Gökçümen and Dr. Matthew Dulik, and extends to them his heartfelt thanks for their invaluable assistance regarding molecular analyses and statistical calculations. Finally, this research could not have been completed without years of guidance, academic investment, and unwavering support of my advisors Dr. Theodore Schurr and Dr. Eduardo Fernandez-Duque.

ABSTRACT

MOLECULAR EVOLUTION OF VASOPRESSIN AND OXYTOCIN RECEPTOR GENES IN OWL MONKEYS (*AOTUS AZARAI*) OF NORTHERN ARGENTINA

Paul L. Babb

Dr. Theodore G. Schurr

The arginine vasopressin (AVP) and oxytocin (OT) hormone pathways are involved in a multitude of physiological processes, and their receptor genes (*AVPR1A* and *OXTR*) have been implicated in increased partner preference and pair bonding behavior in mammalian lineages. This observation is of considerable importance for understanding social monogamy in primates, which is present in only a small subset of primate taxa, including Azara's owl monkeys (*Aotus azarai*). Thus, it is the goal of this dissertation to examine the molecular evolution of *AVPR1A* and *OXTR* in owl monkeys to better understand how the pro-social behaviors related to those loci may have evolved.

However, in order to properly contextualize functional neurogenetic variation related to such sociobehavioral pattterns, it is necessary to first establish the range of molecular variation occurring at non-related genetic loci. To address this issue, we sequenced the entire mitochondrial genome of two species of *Aotus* (*A. azarai* and *A. nancymaae*), and analyzed 39 haplotypes of the mitochondrial COII gene from ten different owl monkey taxa. Next, to understand the recent evolutionary history and genetic structure of our focal owl monkey population, we assessed variation of the mtDNA control region (D-loop) in 118 wild individuals. Furthermore, to establish our knowledge of genetic kinship and individual identity within the wild population, we investigated autosomal variation in the form of 24 short tandem repeat (STR) microsatellite loci.

In concert with these analyses, we characterized the molecular features of *AVPR1A* and *OXTR* in *A. azarai* and other platyrrhines through direct sequencing, and demonstrated that there are substantial sequence differences at both loci across primate species. These data provide new clues on the possible basis of pair bonding in New World species, and may help to explain the sporadic appearance of monogamy in this infraorder. Specifically, despite a common molecular origin, we argue that the AVP and OT pathways have evolved in markedly different ways, due in part to their chromosomal locations and their relative proximity to regions of molecular instability. This study reinforces the notion that neurogenetic loci in primates have undergone significant evolutionary changes, and suggests that monogamy has arisen multiple times in the primate order through different molecular mechanisms.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Chapter 1: Introduction

Summary

Arising from one or several small colonization events roughly 35 million years ago (mya), the New World primates (members of the infraorder *Platyrrhini*) have developed a suite of behavioral patterns that differentiate them from their Old World primate relatives. These patterns include higher degrees of paternal care and monogamous sociality than is reported in other phylogenetic clades of primate taxa. These differences raise the important question: How did such sociobehavioral tendencies develop?

In primates, like other species within the Kingdom Animalia, the evolution of complex behaviors has been related to the evolution of the brain. The evolution of primate social behaviors is no different, as these behavioral patterns find their biological basis in neuronal cell-signaling pathways and the neuroendocrine system. Although influenced by external stimuli, these systems are mediated by endogenous chemicals (and their receptors) that are encoded by genes. Thus, molecular changes that alter the structure or expression of such genes can functionally affect the behaviors they influence.

Interestingly, while some patterns of social behavior fit sytematically into discrete phylogenetic clades (i.e. shared ancestral behaviors), others appear suddenly (evolutionarily speaking) or differ significantly between closely related organisms. This observation implies that some degree of selection must be involved in determining which social behaviors are expressed by individuals within a population of organisms. Social behaviors that positively impact the reproductive fitness of individual actors will be adaptive and spread within the population. Over time, as the sociobehavioral phenotypes

of different populations are selected in different ways, the genetic loci influencing those behavioral phenotypes will exhibit the effects of selection in terms of their sequence variation and patterns of expression.

Evidence for the selective history of specific neurogenetic loci and their associated sociobehavioral phenotypes allows for the development of hypotheses that focus on the ecological, social, or even pathological factors that may have shaped the evolution of those behaviors. When considered together, these elements allow for theorization and the formation of hypotheses of what social behavior might have looked like in the very distant past. In this way, even complex entities such as human societies, whose roots lie in the diverse social behaviors of its members, can be understood in an evolutionary context.

Working under the assumptions that the colonization of the New World by proto-platyrrhines can be roughly pinpointed in time, and that all platyrrhine lineages coalesce to a few hypothetical ancestral genomes, it may be possible to identify neurogenetic differences that relate to the evolution of lineage-specific social behaviors. In addition, the examination of intraspecific genetic variation in a population of primates allows us to estimate rates of mutation and sequence evolution in real time. Thus, by studying neurogenetic diversity of platyrrhines at both micro- and macro-evolutionary levels, we can begin to develop a framework for understanding the broader molecular mechanisms involved in the evolution of plathrrhine social behavior.

1.1 Platyrrhine Evolution

The primate Order can trace its phylogenetic roots back to the Cretaceous-Tertiary boundary over 65 million years ago (mya), when the first primitive mammals were able to survive the cataclysmic event that led to the extinction of all dinosaur species (and countless others) (Fleagle, 1999; Hartwig, 2007). The subsequent evolution of primate taxa is characterized by a series of splits, and has been driven by ecological selective pressures, geological processes, geographic obstacles and even stochastic events. The result is an order whose members exhibit a variety of phenotypic traits in peculiar frequencies.

The Platyrrhini, or New World monkeys, represent an enigmatic evolutionary clade within the primate order. Many hypotheses have been suggested regarding their phylogeographic origins and adaptive radiations within a wide range of ecological niches (Schneider, 2000; Hodgson et al., 2009; Rosenberger et al., 2009). Still, neotropical primate researchers disagree about many details concerning the exact nature of the evolutionary history of the platyrrhines.

Even so, the paleohistory of the Americas and the creation of their diverse range of ecological biomes provided the context and selective pressures for platyrrhine evolution. Hence, in order to study platyrrhine evolution, the history of the Americas must be taken into account. This particular story begins with the arrangement of the continents before the Eocene, as there were no true primates in South and Central America before that period of time (Poux et al., 2006).

Between the mid-Carboniferous and Jurassic periods, 300-200 mya, the familiar continents of today were united in the supercontinent known as Pangaea. By the middle

of the Jurassic, Pangaea had split apart into smaller supercontinents, one of them being Gondwana[1]. Around 150 mya, Gondwana began to split apart, and at approximately 100 mya a violent rift created the Atlantic Ocean between Africa and South America (Tarling, 1980; Houle, 1999; Poux et al., 2006).

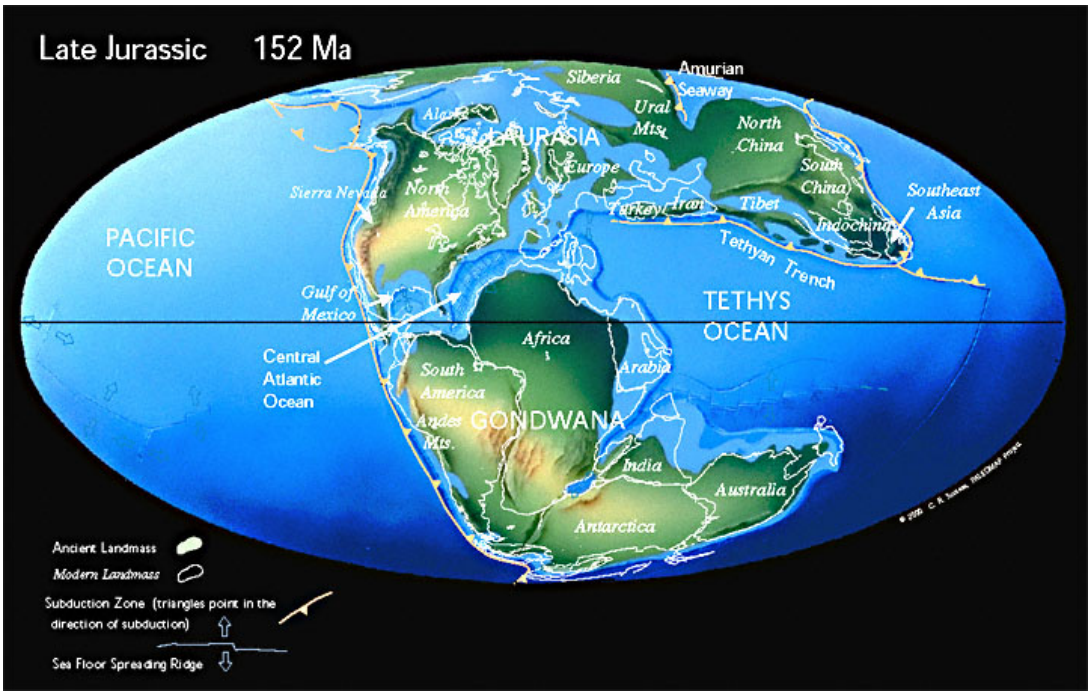


Figure 1.1. Map series depicting the breakup of Pangea into the Laurasian and Gondwanan supercontinents in the Late Jurassic, roughly 150 mya. (Image source: Paleomap Project, Dr. Chris Scotese, http://www.scotese.com/late1.htm).

From this moment in history, for the next 97 million years, South America was isolated from other continents until it reconnected with North America via the Isthmus of Panama ~3 mya (Tarling, 1980). This history implies that any terrestrial organism that lived and evolved in South America should have been on the continent prior to its breakup with Gondwana, or arrived after its reconnection with North America (Poux et

[1] Gondwana is believed to have consisted of the modern day continents of Antarctica, South America, Africa, and Australia, as well as the landmasses of Madagascar, the Arabian Peninsula, and the Indian subcontinent (Tarling, 1980).

al., 2006). Interestingly, the fossil record indicates that platyrrhine ancestors were living on South America as far back as 35 mya. How, then, did platyrrhine primates suddenly appear in South America, and from whence did they come?

Several scenarios for the arrival of platyrrhines to South America have been suggested, intensely debated, and tested against the ever-growing body evidence from fossil, genetic, physiological and geological records. Of these, an African origin for platyrrhine ancestors is regarded as the most parsimonious scenario, as it only requires migration from Africa across the newly formed Atlantic Ocean to South America (Houle et al., 1999). However, lacking the ability to swim, it is not entirely clear how African proto-platyrrhines could have embarked upon such a voyage. It has been suggested that these small primates may have travelled across the ocean on floating patches of vegetation (Lavocat, 1993).

However first initiated, the radiation of proto-platyrrhine populations after their colonization of South America resulted in the creation of many new primate lineages that adapted to a diverse array of ecological niches across the continent (Hugot, 1998; Hartwig, 2007; Hodgson et al., 2009). From a phylogenetic point of view, this is fascinating, as all platyrrhine lineages can hypothetically be traced back to a very few early proto-platyrrhines, with little or no additional gene flow from ancestral populations in Africa. Although, in theory, it should be simple to generate a scenario to explain platyrrhine evolutionary history, the development of an agreed upon phylogeny for the clade has faced many challenges.

It has been difficult for researchers to understand the phylogenetic relationships between primate taxa because different morphological, behavioral and genetic analyses

of platyrrhine traits have classified lineages into different groupings that are highly dependent on the type of comparison being made (Hugot, 1998). There is a general consensus concerning the 16 genera[2] that comprise the platyrrhines. However, the number of major family clades[3] within the infraorder, and the genus membership within each of those three clades, have been hotly contested, and new possibilities for generic classifications have been regularly suggested (Horovitz & Meyer, 1995; Schneider, 2000; Goodman et al., 1998; Hugot, 1998; von Dornum & Ruvolo, 1999; Ray et al., 2005; Hodgson et al., 2009).

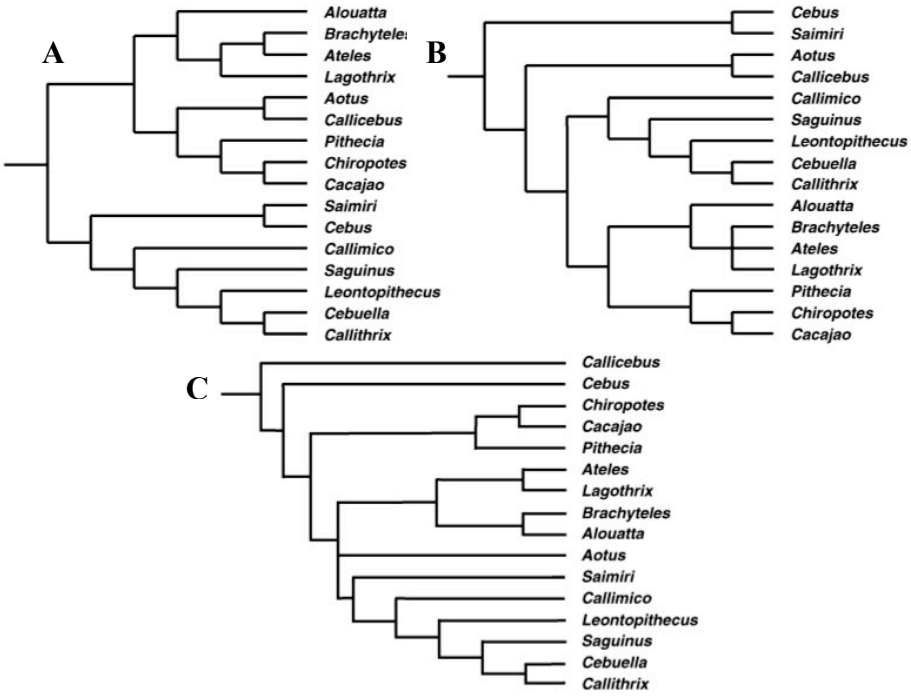


Figure 1.2. Phylogenetic arrangements of platyrrhine primates based on morphologic studies according to **Panel a.**, Rosenberger (1984); **Panel b.** Ford (1986); **Panel c.**, Kay (1990).

[2] *Callithrix*, *Cebuella*, *Leontopithecus*, *Saguinus*, *Callimico*, *Cebus*, *Saimiri*, *Aotus*, *Pithecia*, *Chiropotes*, *Cacajao*, *Callicebus*, *Ateles*, *Lagothrix*, *Alouatta*, and *Brachyteles*

[3] Pitheciidae, Atelidae, Cebidae are usually recognized, wherease Callitrichidae and Aotidae are more recent cladistics distinctions.

Regardless of the generic classification of taxa within the infraorder, platyrrhines can be distinguished from other primates in that they possess widely spaced lateral-facing round nostrils on a broad, flat face (*Platyrrhini* is Latinate for "flat-nosed"), and a dental formula with three premolars per quadrant as a marked ancestral trait (Fleagle, 1999; Hartwig, 2007). By contrast, catarrhines have slanted and downward facing slit-like nostrils, more pronounced noses, and the derived possession of only two premolars per dental quadrant. Yet, despite such morphological differences, it is the patterns of social behavior exhibited by platyrrhine taxa that most distinguish them from their catarrhine relatives.

While the primate order is well known for its diversity of social structures and reproductive strategies, this is particularly true of the platyrrhine primates. Over the last 35 mya, every type of mating system has evolved within the collection of extant platyrrhine taxa (Di Fiore & Rendall, 1994; Rendall & Di Fiore, 2007). In particular, monogamous mating systems and prolonged pair bonds are particularly common among the New World monkeys when compared to their Old World primate relatives. But how did these sociobehavioral tendencies develop?

1.2 Biological Basis of Social Behavior

For over a century, scientists have made voluminous observations about the heritable nature of behavior as a phenotypic 'trait', leading to the general agreement that genes must therefore play some role in creating or directing behavior (Robinson et al., 2005; Stockinger et al., 2005; Burmeister et al., 2008). As genes are subject to mutational change, all behavior, from simple to complex, must be susceptible to change

(and selection), as well (Wilson, 1975; Robinson et al., 2005). The last few decades have provided several novel avenues for addressing these issues through experimental work.

Defining behavior is no easy task. Traditionally, "'behavior' encompasses any orderly movement with recognizable and repeatable patterns of activity produced by members of a species" (Baker et al., 2001:13). These movements or patterns of activity are often viewed as fitting somewhere along a conceptual spectrum, with innate/instinctual behavior at one end and learned/experienced behavior at the other. In this context, where do genes fit into this model? According to Baker and colleagues (2001, 14),

> [a]ny behavior requires the functioning of a multicellular circuit beginning with input to the nervous system, propagation and interpretation of that input in the CNS [Central Nervous System], and output via neurons that directs a response via neuromuscular, or neuroendocrine systems, or both. Impairment of any part of such a circuit is likely to cause decrements in the behavior it subserves.

Genes encode the many parts and pathways that make up the brain, the CNS, and the neuroendocrine system. However, linking an ephemeral behavioral event to a chemical sequence or process can be a hypothetical leap of faith if not properly connected. Behavior itself is complex; it involves a tremendous number of different physiological pathways, many of which have multiple roles within an organism, including the production of other distinct behaviors. Therefore, like any other complex genetic trait, behavior must be reduced to its simpler components.

Sir Francis Galton hypothesized (as noted in Visscher et al., 2008) that all biological behaviors must have a heritable component. More specifically, if it is to be heritable, behavior must be reproducible in successive generations of organisms. Consider the instinctive retrieval behavior of a yellow Labrador retriever versus the

ceaseless pursuing focus of a beagle - two extremes of a similar behavior (chase, track) within the domesticated dog breeds (*Canis lupus familiaris*). In humans, this quality of behavior can be highlighted by familial concordance behavioral tendencies or pathologies, such as the aggregation of mental illness among relatives.

Behavior also has an evolutionary component. By comparing related taxa, such as humans and chimps, rats and mice, or dogs and wolves, we can see a clustering of behavioral similarities. Humans and chimps share behaviors that are characteristic of highly social primates, including nurturing, cooperation, and altruism, even some facial expressions (Pollick & de Waal, 2007; de Waal, 2008). Such similarities also tend to diminish as evolutionary distance between the species increases, while the greatest behavioral similarities exist within populations of the same biological species.

However, it should also be noted that many behaviors often have an exogenous component. The reproducible actions that constitute a particular "behavior" can change their outward expression in response to outside stimuli, or to dramatic endogeneous alterations in biological structures or processes. Numerous examples of these alterations can be drawn from studies in the fields of neuropsychopharmacology and psychiatry, which show that human behavior can be specifically targeted and altered with certain pharmaceutical drugs (Abrahams & Geschwind, 2008; Burmeister et al., 2008). Similarly, when the structure of the brain is damaged through injury or disease (e.g., Alzheimer's and other neurodegenerative diseases), severe behavioral changes can result (Burmeister et al., 2008; Visscher et al., 2008). Depending on the affected region, this damage can be manifested in a number of ways, from slight mood swings to complete catatonic states.

Nevertheless, due to their multifaceted yet reproducible forms, all behaviors, simple or complex, must have some genetic basis. For example, in a series of investigations on the courtship behavior of male fruit flies (*Drosophila melanogaster)*, one of the genes underlying such behavior was empirically revealed. Furthermore, by mapping the neural circuitry related to that locus, molecular biologists were able to confirm the definitive role of genetics underlying simple behavioral components of more complex behaviors (Stockinger et al., 2005).

With its fast life cycle, speedy rates of reproduction, easily recognizable morphological and behavioral phenotypes, and small karyotype (4 pairs of chromosomes), *Drosophila* is a fantastic model organism for genetic manipulation in the laboratory. Its utility is apparent in the realm of behavioral genetics, because, in fruit flies and other invertebrates, "the relatively straightforward connection between neural and behavioral events has enabled us [researchers] to divine the otherwise obscure functional significance of many neural processes and mechanisms" (Edwards et al., 1999, 160). Theoretically, then, these processes and mechanisms will themselves be constituted by specific cellular (neuronal) structures and molecules that can ultimately be traced back to the sequence and expression of specific genes. One *Drosophila* gene in particular, dubbed *fruitless* or *fru*, codes for a specific type of neuronal protein that is at the core of the molecular pathway involved in male courtship (Gill, 1963). Normally, this male courtship consists of a fixed action pattern, the production of which undeniably impacts an individual's reproductive success.

In mutant strains of flies with an altered *fru* gene and disabled *fru* neuronal processing, males display aberrant courtship behavior. Such behaviors include single-

pair courtship, chaining, and performed poorly in competitive mating, fertility, receptivity and egg laying assays (Demir & Dickson, 2005). Mutant males performed as well as control males (wild-type) in assays for locomotion, flight, phototaxis, odor avoidance, taste sensitivity, and taste discrimination. Hence, the *fru* neurons function directly in male courtship and, indeed, are largely dedicated to this behavior (Stockinger et al., 2005).

While *fru* proteins are visibly expressed in many the projections of neurons, regardless of sex, they are particularly enriched in the sensory neurons, with special attention paid to the olfactory receptor neurons (ORNs). Through a sophisticated series of molecular manipulations, investigators were able to show that the *fru* gene, *fru*-protein bearing neurons, *fru* ORNs and external receptors are interconnected in an olfactory circuit dedicated to the detection and processing of volatile sex pheromones (Stockinger et al., 2005). By altering the genetic structure of just one component, the *fru* gene, they were able to change the whole system, resulting in completely different behavioral performances by male files. These robust findings link a distinct behavior to a genetic basis and neurochemical pathway.

1.3 Selection of Sociobehavioral Traits

Throughout evolutionary history, behaviors have grown more "complex". They are no longer simply reactive responses to the presence of a particular stimulus, but instead processes involving the recognition or continuation of relationships after an initial interaction. All of these behavioral modalities (movement, sensory perception, reactivity, conditioning and cognition) can be seen as adaptations to new evolutionary challenges

and environmental contexts (Skinner, 1984). Complex behaviors involved in sociality, or the interactions between organisms of the same population or species, can similarly be hypothesized to have developed as adaptive traits (Wilson, 1975; Robinson et al., 2005). However, social behaviors must additionally negotiate the balance between cooperation and competition with other organisms within a group, presenting the possibility for the differential selection of social behaviors for populations living in different ecological and social settings (Trivers, 1972; Wilson, 1975; Cheney & Seyfarth, 1999; Robinson et al., 2005). A scenario like this might explain the wide range of diverse social behaviors observed across the Animal Kingdom.

Primates have evolved an incredible array of behaviors that maximize fitness through social living. Macaques can easily differentiate kin from non-kin, as well as keep running tabs of dominance ranking along the matrilineal lines (Leonard et al., 1985; MacKenzie et al., 1985). Female baboons trace the hierarchical structures existing within their troop, and form "friendships" with non-alpha males to protect un-weaned offspring from potentially infanticidal new alpha males (Cheney et al., 1996; Cheney & Seyfarth, 1999; Palombit, 1999). Chimpanzees often utilize forms of deception or acts of altruism over resources with conspecifics, either to maximize fitness immediately or in the future (Pusey et al., 1997; Pollick & de Waal, 2007; de Waal, 2008). In addition, squirrel monkeys can adapt their social structure and dominance hierarchies in response to changing ecological conditions and resources (Boinski, 1999).

However, no other primate has so effectively manipulated its social surroundings as modern humans. Humans generate and accumulate vast amounts of socially learned knowledge, employ mass communication and complex technologies, follow specific

social rules and rituals, and engage in symbolic thought and language. Indeed, maximizing the benefits of sociality is humanity's evolutionary niche (Alvard, 2003). However, such extraordinary levels of sociobehavioral development in our own species are exceedingly difficult to associate with any molecular mechanism(s), as there are few, if any, other comparable behavioral suites in other organisms to analyze.

Thus, understanding the evolution of social behavior in humans and other primates benefits from the reduction of larger and more complex sociobehavioral elements to simpler, well-defined behavioral phenotypes. Once having defined these sociobehavioral phenotypes, researchers can examine their association with specific genetic loci and identify noteworthy patterns of molecular variation present at their genomic locations (Phelps et al., 2009). Phylogenetic relatives provide natural comparative case studies for molecular variation. Therefore, by assessing molecular variation at these loci, it is possible to reveal transcriptional or structural changes to gene dosage or function that may occur as a result of this variation, as well as enable the identification of signatures of directional selection that might characterize a particular evolutionary lineage.

In this regard, because different lineages of organisms will exhibit different types of social behavior and will possess different patterns of molecular variation, it becomes possible to develop hypotheses about the ecological and social drivers that may have selected for particular sociobehavioral phenotypes based on the empirical analysis of the genes which have helped generate them. By examining genes that code for neurological processes and neuroanotomical structures involved in social behavior, we are increasingly able to examine their selective history at the molecular level, and to

potentially generate an empirical measure of evolution for a particular complex behavioral phenotype (Robinson et al., 2005; Carter et al., 2008).

1.4 Neurogenetic Evolution

Genes encode every molecule, transporter, regulator, and mechanism involved in the workings of multi-cellular circuits. Thus, any modification of a coding or regulatory sequence through random mutation can potentially have profound effects on the downstream functioning of the circuit in which that particular locus operates (Baker et al., 2001; Lim et al., 2004). If enough of these alterations accumulated in a neural circuit, then the behavior exhibited by an organism possessing these mutations could be changed as a result. And, if these neurogenetic changes increased in frequency in a population over time, then the resultant change in behavior could even lead the population down a novel evolutionary or adaptive path (Enard & Pääbo, 2004; Robinson et al., 2005). Thus, through the identification of functional neurogenetic loci, researchers now possess the ability to identify proximate molecular causes for the variation in phenotypic behavior, and estimate how those behaviors may have evolved.

Recent genetic studies of primate and mammalian species have shed light on the factors influencing the evolution of behavior in different taxa. Research conducted on the serotonin transporter gene in rhesus macaques (*Macaca mulatta*) has shown that molecular variants in the serotonergic pathway (*5-HT* and *SERT*) resulting from a short tandem repeat (STR) insertion within the promoter region can greatly affect levels of aggression exhibited by different individuals (Lesch et al., 1996; Champoux et al., 2002). This pattern mirrors those seen in many other studies that have focused on the

functioning of serotonin in humans, the pathways of which are major concerns of psychiatric and pharmaceutical research (Sabol et al., 1998). Similarly, the elucidation of the related monoamine oxidase A (MAOA) neurogenetic pathway revealed that repeat promoter polymorphisms cause variable aggressive tendencies in macaques, and that this association correlates with patterns of impulsivity, natal dispersal and maternal rejection (Bennett et al., 2002; Newman et al., 2005, Maestripieri et al., 2006a, b). Additional work has also reported variation in the neurogenetic expression levels of dopamine (DA) and catechol-O-methyltransferase (COMT) due to promoter alteration leading to changes in transcriptional regulation (Tenhunen et al., 1994; Palmatier et al., 1999).

Several other candidate pathways, molecules and hormones have been implicated in the evolution of pro-social behaviors in mammals, with prolactin, oxytocin, vasopressin, testosterone and dopamine topping the list. Two neuroendocrine pathways in particular, arginine vasopressin (AVP) and oxytocin (OT) have been implicated in the maintenance of social distance and tolerance (Carter, 1998; Hammock & Young, 2004, 2005; Young et al., 2005; Carter et al., 2008). By studying two genes coding for signal receptors within these pathways, *AVPR1A* (AVP pathway) and *OXTR* (OT pathway), and the inter- and intraspecific genetic variation that occurs within and around them, we are afforded an excellent case study for analyzing molecules that may regulate the occurrence of pro-social behaviors and the overall maintenance of social interactions. Data from this research lays the groundwork for understanding the development of the various social systems that occur across primate species.

Table 1.1. Social behaviors of various organisms and related genetic loci.

Behavior	Organism	Gene	Molecular function
Mate recognition and courtship			
Vocal learning, vocalization	*Taeniopygia gutta; Homo sapiens*	*FOXP2*	Transcription factor
Vocal learning, song recognition	*T. guttata*	*zenk (Zif269/Egr1/ NGFIA/Krox24)*	Transcription factor
Pheromone-mediated communication	*Mus musculus*	Vomeronasal *V1R, V2R*	G-protein receptors
Pheromone-mediated communication	*D. melanogaster*	*Gr68a (Gustatory receptor 68a)*	G-protein receptor
Pheromone-mediated communication	*Bombyx mori*	*BmOR1 (olfactory receptor 1)*	G-protein receptor
Male courtship	*D. melanogaster*	*fruitless*	Transcription factor
Male courtship; timing of mating	*D. melanogaster*	*period*	Transcription cofactor
Female receptiveness (lordosis)	Rodents	Oestrogen responsive genes	Various functions
Post-mating behaviour			
Refractoriness to mate, ovipositioning	*D. melanogaster*	Genes for seminal proteins	Various functions
Monogamy, paternal care	Rodents	*V1aR (AVPR1A), OTR (OXTR)*	G-protein receptors
Maternal care	*Rattus norvegicus*	*GR*	Glucocorticoid receptors
Attachment to mother	*Mus musculus*	*Orpm (μ)*	Opioid receptor
Maternal care, pup retrieval	*Mus musculus*	*Dbh (dopamine â-hydroxylase)*	Biosynthesis of norepinephrine
Social hierarchies			
Territorial versus non-territorial males	*Haplochromis burtoni*	*GnRH1*	Gonadotropin releasing hormone
Dominant versus subordinate males	*Procambarus clarkii*	*5HTR1, 5HTR2*	Serotonin receptor
Dominance interactions			
Aggression	*Mus musculus*	*MAO-A*	Monoamine oxidase
Aggression	*Macaca mulatta*	*5HTT*	Serotonin transporter
Subordinate behavior	*Mus musculus*	*Dvl1*	Wnt-receptor signalling pathway

Table adapted from Robinson et al., 2005.

Because they both code for rhodopsin-like G-protein coupled receptor (GPCR) membrane proteins, *AVPR1A* and *OXTR* possess structural similarities to one another as well as other GPCRs that constitute sensory perception hardware (e.g. rhodopsin light receptors in the eye, olfactory odorant receptors, etc.) in other cellular systems. It is easy to imagine a scenario where the AVP and OT neuroendocrine hormone pathways have evolved from more "primitive" sensory perception pathways (Hammock & Young, 2004, 2005; Lim et al. 2004; Carter et al., 2008). Genomic scans of six mammalian genomes have shown that sensory perception is a primary ontological category of genes that has

undergone directional selection in mammalian evolution (Kosiol et al. 2008). Perhaps neuroendocrine GPCRs have been directionally selected, as well.

It is possible that some version of an ancestral gene, the function of which was to perceive the presence of environmental objects or phenomena, took on new adaptive roles in certain organisms, allowing their perception of other social organisms. Those organisms that possessed the modified versions were better able to navigate their social environment, resulting in the promotion of their fitness and the selection of their particular sociobehavioral phenotype in future generations. In this way, even a small mutational change could have had significant functional consequences for the evolution of sociobehavioral capabilities.

1.5 Neurogenetic Variation and Platyrrhine Sociobehavioral Evolution

In reality, "sociality" does not exist as a single behavior. Even "pair bonding" is an exceedingly complex behavioral trait in its various manifestations. Social behaviors occur in a spectrum of intensity when assessed across mating systems of different animals, including primates. Picture the diversity of social behaviors that include a huddled group of pair-bonded titi monkeys twining tails with their offspring, the nepotistic hierarchies of baboons and macaques, or the solitary wanderings of male orangutans (MacKenzie et al., 1985; Cheney & Seyfarth, 1999; Mayeaux et al., 2002; de Waal, 2008). This variation in primate social behavior is likely the result of shifting the balance of cooperation and competition between social actors in the context of different ecological factors (Trivers, 1972; Skinner, 1984). Thus, understanding the mechanisms

that allow social actors to approach, sense, and cognitively recognize other social actors is vital to elucidating the evolution of sociality.

Thus, by defining the social behaviors of a species and reducing them to measurable phenotypic components, one can construct hypothetical scenarios of causal candidate genetic loci that are associated with those components. This kind of approach may enable researchers to trace the molecular history of the genetic regions (via sequence and expression comparisons across or taxa) and reconstruct the evolution of the behavior under study (Enard & Pääbo, 2004; Phelps et al., 2009). Similarly, the directionality of selection can be inferred from patterns of sequence divergence in the candidate genetic loci, and used to generate explanations for the evolution of these loci and behaviors. The data may also provide the necessary context to identify the selective forces (ecological, social, pathological) involved in the development and maintenance of complex social behaviors like mate selection.

But how did these evolutionary events take place, and when did they occur in mammalian history? Do the related pathways function in different ways in closely related phylogenetic lineages? What genetic changes had to be occur to create such specific sociobehavioral pathways from pre-existing genetic machinery?

The evolution of the neurogenetic components underlying social behavior must have had its origin in changes to the structure and expression of genes that significantly affected their function (Enard & Pääbo, 2004). Most behavioral phenotypes are polygenic in origin, and thus provide a large genomic target for the accumulation of molecular variation or recombination events. Single nucleotide polymorphisms (SNPs) and other small mutations could alter the product of genes through novel, gain-of-

function mutations, while mutations within the regulatory regions or the presence of genomic copy number variations (CNVs) could likewise create new expression patterns (Perry et al., 2007; Sebat et al., 2007; Lee et al., 2008). When viewed together, SNPs and CNVs may have provided the necessary mechanisms for both slow- and fast-paced adaptive evolution, ultimately leading to the selection of a wide range of social behavioral phenotypes. Through the detection of these molecular changes, researchers have begun to uncover the genetic underpinnings of some of the behavioral elements that lead to the manifestation of complex behavioral phenomena such as the evolution of social systems in primates (Di Fiore, 2003).

This study represents an attempt to explore questions regarding the evolution of pro-social behaviors through the analyses of neurogenetic variation in *AVPR1A* and *OXTR* in a wild population Azara's owl monkeys (*Aotus azarai azarai*), and other among platyrrhines. Azara's owl monkeys exhibit social behaviors such as the formation of pair bonds and significant paternal care to offspring. Because these behavioral phenotypes are exceedingly rare among the majority of primate taxa, owl monkeys serve as a useful model for investigating the relationship between molecular variation at these neurogeneic loci and the social behaviors (and mating systems) exhibited by different lineages of primates. Thus, it is the goal of this study to examine the molecular evolution of *AVPR1A* and *OXTR* in owl monkeys to better understand how the pro-social behaviors related to those loci may have evolved.

Chapter 2: MtDNA diversity in Azara's owl monkeys

Summary[4]

Owl monkeys (*Aotus spp*.) inhabit much of South America, yet represent an enigmatic evolutionary branch among primates. While morphological, cytogenetic, and immunological evidence suggest that owl monkey populations have undergone isolation and diversification since their emergence, problems with adjacent species ranges and sample provenance have complicated efforts to characterize genetic variation within the genus. As a result, the phylogeographic history of owl monkey species and subspecies remains unclear, and the extent of genetic diversity at the population level is unknown.

To explore these issues, we analyzed mitochondrial DNA variation in our study population of wild Azara's owl monkeys (*Aotus azarai azarai*) living in the Gran Chaco region of Argentina. We sequenced the complete mitochondrial genome from one individual (16,585 base pairs (bp)) and analyzed 1099 bp of the hypervariable control region (CR) and 696 bp of the cytochrome oxidase II (COII) gene in 117 others. In addition, we sequenced the mitochondrial genome (16,472 bp) of one Nancy Ma's owl monkey (*A. nancymaae*). Based on the whole mtDNA and COII data, we observed an ancient phylogeographic discontinuity among *Aotus* species living north, south, and west of the Amazon River that began more than eight million years ago. Our population analyses identified three major CR lineages, and detected a high level of haplotypic diversity within *A. a. azarai*. These data point to a recent expansion of Azara's owl

[4] **The results presented in this chapter have been published in the peer-reviewed report:**
Babb PL, Fernandez-Duque E, Baiduc CA, Gagneux P, Evans S, Schurr TG. (2011). MtDNA diversity in Azara's owl monkeys (*Aotus azarai azarai*) of the Argentinean Chaco. *American Journal of Physical Anthropology 146*, 209-224.

monkeys into the Argentinean Chaco. Overall, we provide a detailed view of owl monkey mtDNA variation at genus, species and population levels.

2.1 Introduction

The South American Gran Chaco is comprised of 1,000,000 km^2 of grassland and forests found throughout Argentina, Bolivia, Brazil, and Paraguay. It extends 1,500 km from north to south, and 700 km from east to west (18º-35º S, 57º-66º W, de la Balze et al., 2003). Following the Amazonian rain forest, the Gran Chaco is the second largest biome of the continent (Bertonatti and Corcuera, 2000), yet, its ecological development and paleohistory are poorly understood.

Chacoan fauna are characterized by high diversity and low endemism (Porzecanski and Cracraft, 2005). Among the inhabitants of the Chaco are some of the southernmost primates in South America, Azara's owl monkeys (*Aotus azarai azarai*), which may have achieved their present day locations via southward migrations along the Paraná-Paraguay Rivers (Zunino et al., 1985). Other Chacoan primate species may have originated from Amazonian stocks to the north, and eventually populated the region through the continent's waterway corridors (e.g., black howler monkeys, *Alouatta caraya*: Do Nascimento et al., 2007; Zunino et al., 2007). However, little is known about the timing of these Chacoan migrations, or how they may have shaped the genetic diversity in southern owl monkey populations.

Questions concerning owl monkey origins extend to the entire genus. In fact, researchers have only recently begun to agree on the number of extant *Aotus* species (Ford, 1994; Defler and Bueno, 2007; Fernandez-Duque, 2011). When initially

described, the genus included only the species *Aotus trivirgatus* (Brumback et al., 1971), although further cytogenetic characterization revealed that this taxon had at least three different chromosomal backgrounds (2n = 46-58) (e.g., Brumback, 1973, 1974; Ma, 1981, 1983). This diversity led to the current designation of thirteen owl monkey species and subspecies based on karyotypes, pelage coloration, and relative levels of susceptibility to different malaria pathogens (*Plasmodium spp.*) (Hershkovitz, 1983; Ford, 1994; Defler and Bueno, 2003; Di Fiore et al., 2009; **Table 2.1**). However, issues such as the close proximity of species' ranges to one another, hybridism, questionable sample provenance, and the difficulties of tracking nocturnal arboreal primates has meant that few samples have been used to characterize the range of genetic variation within the genus (Ashley and Vaughn, 1995; Defler and Bueno, 2007; Plautz et al., 2009; Menezes et al., 2010, Monsalve and Defler, 2011). This is particularly true with regard to the southernmost owl monkey species *A. azarai*.

Table 2.1. Descriptive features of putative species and subspecies of *Aotus*.

Species	Geographic range	Karyotype		Malarial Resistance (*P. brasilianum*)	Coat coloration
		Male	Female		
Aotus azarai azarai	Argentina, Paraguay, Bolivia	49	50	No	Red
Aotus azarai boliviensis	Bolivia	49	50	No	Red
Aotus infulatus	Brazil	49	50	No	Red
Aotus lemurinus	Panamá, Colombia, Venezuela	55	56	Yes	Grey
Aotus lemurinus brumbacki	Colombia	50	50	Yes	Grey
Aotus lemurinus griseimembra	Panamá, Colombia, Venezuela	52-54	52-54	Yes	Grey
Aotus miconax	Perú	Unk	Unk	Unk	Red
Aotus nancymaae	Brazil, Perú	54	54	Yes	Red
Aotus nigriceps	Brazil, Perú	51	52	No	Red
Aotus trivirgatus	Colombia, Venezuela, Brazil, Guyana	Unk	Unk	Yes	Grey
Aotus vociferans	Colombia, Brazil, Perú, Ecuador	46-48	46-48	Yes	Grey
Aotus lemurinus zonalis	Panamá	Unk	Unk	Unk	Grey
Aotus hershkovitzi	Colombia	58*	58*	Unk	Grey

Data Sources: Brumback, et al. 1971; Brumback, 1973, 1974; Defler and Bueno, 2003, 2007; Defler et al., 2001; Hershkovitz, 1983; Ma et al., 1976a, b, 1977, 1978, 1985; Ma, 1981, 1983; Ford, 1994; Nino-Vasquez et al., 2000.
* Close to *A. lemurinus* or *A. vociferans*, only noted from four specimens at one locale. Unk = Unknown

In addition, phylogeographic studies of the genus have estimated molecular divergence dates that are not consistent with fossil and cytogenetic evidence. For example, estimates of 3.6 mya (Ashley and Vaughn, 1995) or 4.7 mya (Plautz et al., 2009) for the divergence of *Aotus* species do not agree with paleontological evidence like the 11.8-13.5 mya *Aotus didensis* fossils from La Venta, Colombia (Setoguchi and Rosenberger, 1987; Rosenberger et al., 2009; Takai et al., 2009). They are also not congruent with coalescence dates of ~22 mya for an *Aotus*-platyrrhine divergence based on nuclear DNA data (Opazo et al., 2006), or ~15 mya for the emergence of the genus based on whole mitochondrial genome sequences (Hodgson et al., 2009). If those estimates (3.6 - 4.7 mya) for the diversification of the genus were accurate, they would imply more than 10 million years of lineage stasis before extant owl monkey species began to diverge from one another.

Given these apparent discrepancies, we were interested in exploring further the evolutionary history of *Aotus* through the analyses of molecular genetic data. We hypothesized that the pattern and timing of the radiation of *Aotus* species within South America was more complex and began earlier than previously postulated. To elucidate the timing and nature of speciation events, and to ascertain the position of our study taxon within the phylogenetic history of *Aotus*, we characterized mtDNA variation in a wild population of *A. a. azarai* living in the Argentinean Gran Chaco. We anticipated that the southern species of owl monkeys would be characterized by the accumulation of commonly derived mutations, reflecting the progressive settlement of individuals into regions south of the Amazon River. By contrast, we predicted that the northern owl monkey taxa would exhibit greater haplotype diversity as a result of barriers to gene flow

caused by geological and hydrological change, or perhaps driven by natural forces selecting for different pathogen regimes.

To test our predictions, we sequenced the entire mitochondrial genome of one Azara's owl monkey (*A. a. azarai*) and one Nancy Ma's owl monkey (*A. nancymaae*). We investigated the phylogenetic origins of *A. a. azarai* by examining the mitochondrial COII gene (Ruvolo et al., 1993; Adkins and Honeycutt, 1994) and compared our data with those from other species within the genus *Aotus*. We further utilized information on the hypervariable CR to characterize the structure of genetic diversity of the study population. By implementing these three approaches, we were able to describe owl monkey mitochondrial evolution at the genus, species, and population levels.

2.2 Methods

Study area

The Owl Monkey Project (*Proyecto Mirikiná*) studies the Azara's owl monkey population that inhabits the gallery forests along the Pilagá and Guaycolec Rivers (Fernandez-Duque et al., 2001), within the province of Formosa, Argentina (**Figure 2.1**). Azara's owl monkeys have a species range that extends at least another 180 km south and 300 km west from the study area (Zunino et al., 1985), and there is no evidence suggesting that the study population has been isolated as a result of its geographic location or human activities.

Figure 2.1. Study area location in Formosa Province, Argentina. Core study site is located at latitude: 25° 59.4' South, longitude: 58°, 11.0' West, and projected using the WGS 1984 coordinate system, geographic panel UTM 21S.

Samples

Since 2001, over 160 individuals have been sampled within the 3 km^2 core study area**.** Upon capture, each animal was given a physical exam during which hair, blood, or tissue samples (ear punches, skin biopsies) were collected for use in genetic analyses (Fernandez-Duque and Rotundo, 2003). For one individual, the source of DNA was the remains of a placenta and fetus found on the ground in the savannah.

From the collected samples, we characterized mtDNA variation in 118 *A. a. azarai* individuals. Ninety-one of the samples were from individuals who inhabit the core study area. Seven other samples were from individuals captured along the gallery forest as far as 10 km upstream from the core area, while 15 individuals were sampled

downstream of this location. One sample came from a male who was captured in the gallery forest along the Monte Lindo River, ~25 km north of the field site. In addition, we collected four samples from captive individuals of unknown geographic origin at the Saenz-Peña Municipal Zoo, located 250 km away from the study area in the city of Saenz-Peña, Chaco Province, Argentina.

For comparative analyses, DNA samples from four putative *Aotus* species and subspecies (*A. nancymaae*, *A. nigriceps*, *A. lemurinus*, and *A. l. grisiemembra*) and six individuals representing three other platyrrhine taxa (two *Callicebus donacophilus*, two *Pithecia pithecia pithecia,* and two *Saimiri sciureus sciureus*) were obtained from the Zoological Society of San Diego. Another five samples of *A. nancymaae* were obtained from individuals at the DuMond Conservancy for Primates and Tropical Forests (Miami, FL).

DNA was isolated from tissue, blood, and hair roots using QIAamp purification kits (Qiagen), and DNA quantity and quality were assayed on the NanoDrop ND-1000 spectrophotometer (Thermo Scientific).

Genetic sequencing

To avoid the amplification of nuclear insertions of mtDNA (numts) in all downstream reactions, the entire mtDNA genome was amplified in two large fragments (LR1 and LR2) for all samples, each ~9 kb in length, with >200 and 1100 bp of overlap between the ends of the fragments (Raaum et al., 2005; Thalmann et al., 2005; Sterner et al., 2006). Using two pairs of primers designed for human mtDNA (Meyer et al., 2007), we conducted Long Range polymerase chain reactions (LR-PCR) with the Expand Long

Range dNTPack (Roche) following the protocol and amplification parameters recommended by the manufacturer.

Next, to determine the unknown sequence of the entire mitochondrial genomes for both *A. a. azarai* and *A. nancymaae*, we designed a panel of twenty overlapping primer pairs to obtain complete mtDNA sequences for samples that had been amplified through LR-PCR. These primers were based on conserved regions shared by *Aotus lemurinus* (FJ785421), *Saguinus oedipus* (FJ785424), and *Aotus trivirgatus* (AY250707). Stretches of consensus (100% shared base identity) extending for ≥20 bases in length were screened for their capacity to function as primers using NetPrimer (Premier BioSoft). Portions of the *A. lemurinus* mitochondrial genome were also directly interrogated using Primer3 (Rozen and Skaletsky, 2000; SourceForge.net). In total, the two methods yielded a pool of 57 pairs of potential primers. After re-alignment with the *A. lemurinus* sequence, twenty pairs of primers were selected to amplify the mtDNA genome in overlapping fragments ranging from 750 to 1800 bp in size (**Figure 2.2, Table 2.2**).

To address questions related to the phylogeographic origin and phylogenetic placement of the study population, we characterized sequence diversity in the COII gene. Using primers redesigned from sequences available in the published literature (Disotell et al., 1992; Ruvolo et al., 1993; Ashley and Vaughn, 1995), we amplified and sequenced the entire COII gene (696 bp in platyrrhines) from LR-PCR products for all samples.

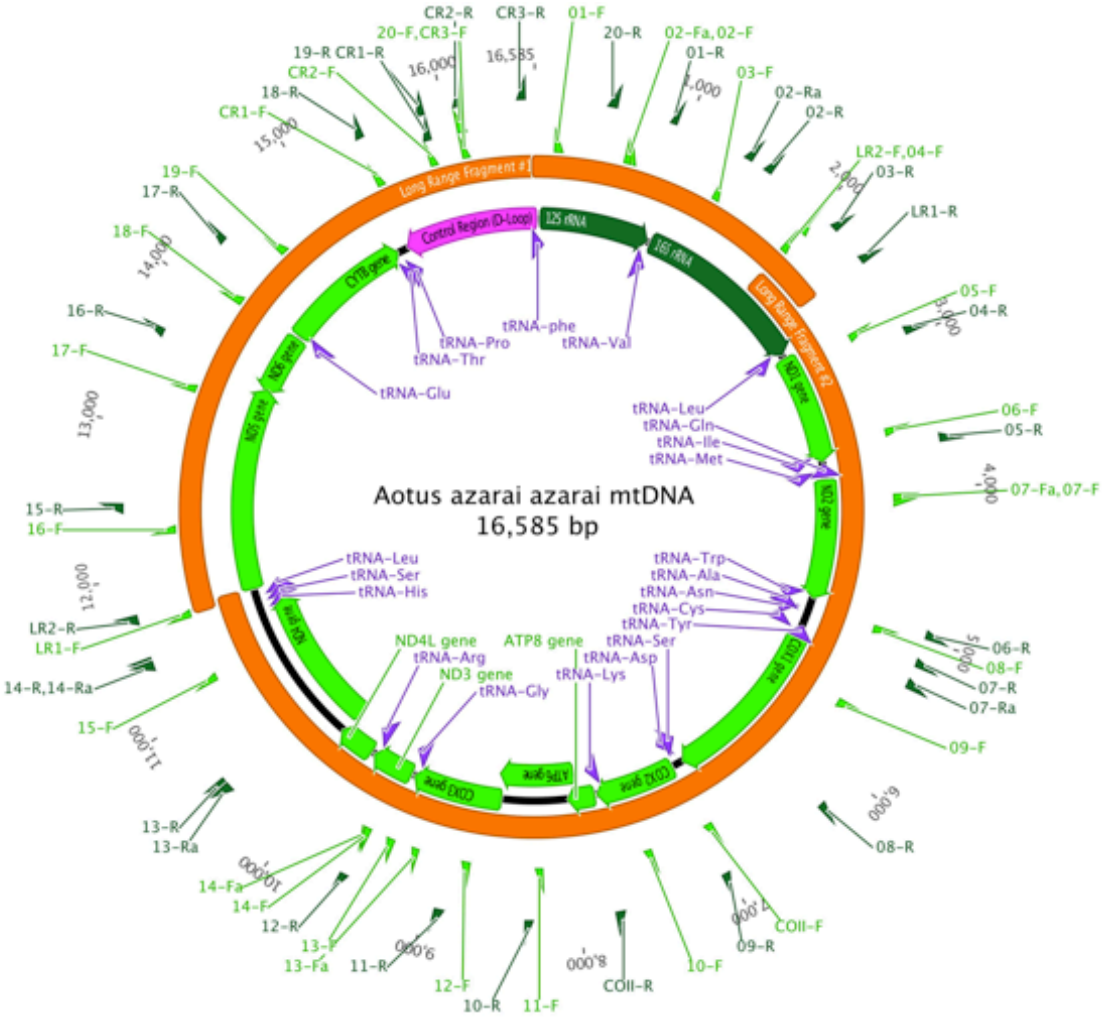


Figure 2.2. Schematic diagram of the mtDNA molecule in *A. a. azarai,* showing the PCR amplification and sequencing strategies used in this study. Arrows indicate the position and direction of the oligonucleotide primers listed in **Table 2.2**. Detailed PCR recipes and cycling parameters listed in **Table 2.3**.

Table 2.2. *Aotus*-specific primers for amplification and sequencing of the mitochondrial CR and COII gene.

Primer name	Genetic region	Oligonucleotide sequence (5'-3')
LR1-F	Long Range Fragment 1	GGC TTT CTC AAC TTT TAA AGG ATA
LR1-R	Long Range Fragment 1	TGT CCT GAT CCA ACA TCG AG
LR2-F	Long Range Fragment 2	CCG TGC AAA GGT AGC ATA ATC
LR2-R	Long Range Fragment 2	TTA CTT TTA TTT GGA GTT GCA CCA
01-F	Whole MtDNA	TGA GGA GCG AGT ATC AAG CAC
01-R	Whole MtDNA	GTG ACG GGC GGT GTG TG
02-F	Whole MtDNA	TCG CAG AGT AAG CAG AAG CA
02-R	Whole MtDNA	CTA TGG TGG TGG AGC GTT TT

Primer name	Genetic region	Oligonucleotide sequence (5'-3')
02-F.2	Whole MtDNA	CGC CAT CTT CAG CAA ACT CC
02-R.2	Whole MtDNA	GGA CAA CCA GCT ATC ACC A
03-F	Whole MtDNA	CTA TGT GGC AAA ATA GTG GG
03-R	Whole MtDNA	CCA TAG GGT CTT CTC GTC TTA
04-F	Whole MtDNA	GGT AGC ATA ATC ACT TGT TCT C
04-R	Whole MtDNA	CTA GGG TTG GGG CAG TTA CA
05-F	Whole MtDNA	CAA TTT CGC AAA GGT CCT AAC
05-R	Whole MtDNA	CTT ATG TTT GGG GTG GAA TGC
06-F	Whole MtDNA	CGA TTC CGA TAC GAC CAA CT
06-R	Whole MtDNA	GGC TTT GAA GGC TCT TGG TC
07-F**	Whole MtDNA	GCA ACC GCA TCC ATA ATT CT
07-R**	Whole MtDNA	CGG CGG GAG AAG TAG ATT G
07-F.2	Whole MtDNA	GCT CCA CAG AAG CAT CCA CT
07-R.2	Whole MtDNA	GAG TAA GCA TTA GAC TGT AAA TC
08-F	Whole MtDNA	GCA TCA ACT GAA CGC AAA TC
08-R	Whole MtDNA	ATG ATT ATA GTG GCT GAT GT
09-F	Whole MtDNA	GCT TCT GAC TTC TAC CCC CAT C
09-R	Whole MtDNA	GGT GTT GCC ATT AAG ATA TA
10-F	Whole MtDNA	CAA CCC TCC CAA TAG AAG CA
10-R	Whole MtDNA	AGT GGG ACA GGT GTT CCT TG
11-F	Whole MtDNA	CTA TGG GCA GCA ACC GTA
11-R	Whole MtDNA	GAG TGG TAG AAT GCT CAG AAG
12-F	Whole MtDNA	CGT TGT CCG AGA GGG TAC AT
12-R	Whole MtDNA	TAA GGG TTG TGT TTT TCG GC
13-F**	Whole MtDNA	CTA TAT CTC TAT CTA CTG ATG AGG
13-R**	Whole MtDNA	GAG TGG GGA TAA GGG TGG TT
13-F.2	Whole MtDNA	TTT CTG ACG GAA TTT ACG GC
13-R.2	Whole MtDNA	CTG TGG CCG TGA ATG TTA TG
14-F**	Whole MtDNA	GCC GAA AAA CAC AAC CCT
14-R**	Whole MtDNA	GTA TGT CAG TGG CCC TCG TT
14-F.2**	Whole MtDNA	GAA TGT GGA TTT GAC CCC AC
14-R.2**	Whole MtDNA	CGT GTG AAT AGG GGT TTT ACA TT
14-F.3	Whole MtDNA	CCT AAC CCT CAC AGC CTG AC
14-R.3	Whole MtDNA	GGC AAG GTT GGC TAG ATT TG
15-F	Whole MtDNA	TAC GAA CGA ATT CAC AGC CG
15-R	Whole MtDNA	ACT GGG GTA GGT CCT TCT ATA GC
16-F	Whole MtDNA	AGC AAT AGC ATG ATT CTT CCT A
16-R	Whole MtDNA	ATT ATG GTG TTT GAG TTG TT
16-F.2	Whole MtDNA	CTC CTT CCC CCT AAT AAG TCT CC
16-R.2	Whole MtDNA	GTT GTT TTG GTT ACT TGT TG
17-F	Whole MtDNA	ACC CCC ACT CAA GCC TAA CT
17-R	Whole MtDNA	GCG GTT GAG GTA TCT GGT GT
17-F.2	Whole MtDNA	ACC AAA ACT AAC AAT ACA AAC TC

Primer name	Genetic region	Oligonucleotide sequence (5'-3')
17-R.2	Whole MtDNA	GGC GAC GGA GGA GAA GGC
18-F	Whole MtDNA	GTC ATT ATT CCC ACA TGG AC
18-R	Whole MtDNA	GGG TCT TGG CTG GTA GTT CA
19-F	Whole MtDNA	GCC AAA TAT CAT TCT GAG G
19-R	Whole MtDNA	CTG GTT TCA CGG AGG TAG GT
19-F.2	Whole MtDNA	GAG GTG GCT TCT CAG TAG
19-R.2	Whole MtDNA	GAA GTG GGC GGG TTG CTG
20-F	Whole MtDNA	CTC AGC ATT CCC GTA GGT TC
20-R	Whole MtDNA	CTA AGC ATA GTG GGG TAT CTA ATC
CR1 F	CR	TGA ACT ACC AGC CAA GAC CC
CR1 R	CR	CTG GTT TCA CGG AGG TAG GT
CR2 F	CR	CTA CCT CCG TGA AAC CAG CA
CR2 R	CR	TAC GGG AAT GCT GAG GAA AG
CR3 F	CR	CTT TCC TCA GCA TTC CC
CR3 R	CR	TTT TCT GAA GGG TGT GGT TT
tRNA-Asp-F [a,b]	COII	AAC CAT TCA TAA CTT TGT CAA
tRNA-Lys-R [a,b]	COII	CTC TTA ATC TTT ACT TAA AAG
COII-Seq [a,b]	COII	TTT AGG CGT CCT GGG ATT
COII-F.2	COII	AAT AAT TAC ATA ACT TTG TCA A
COII-R.2	COII	CTC TCG GTC TTT AAC TTA AAA G
COII-int-F	COII	GGC CAT CAA TGA TAC TGA AGC
COII-int-R	COII	TTC ATA GCT TCA GTA TCA TTG ATG G

Data sources: [a] Ruvolo et al., 1993, [b] Ashley and Vaughn, 1995. All remaining primers designed for current study.

** Primers that failed to work during amplification PCRs. Such primers were re-designed, and are denoted by the suffix –.2 or –.3 at the end of the primer name.

Concurrently, for the analysis of population-level maternal diversity, we targeted the entire hypervariable control region (CR) of the mtDNA molecule. Following the enrichment of mtDNA fragments through LR-PCR, three primer pairs (CR-1 thru CR-3) were used to obtain 1099 bp of CR sequence for each individual.

PCR cycling parameters were optimized for each primer pair using the Touchgene Gradient thermocycler (Techne), and all subsequent reactions were amplified in GeneAmp 9700 thermocyclers (ABI). Amplified products were visualized on 1% TBE SeaKem agarose (Lonza) via gel electrophoresis. PCR recipes and parameters are

detailed in **Table 2.3**.

Table 2.3. Parameters for amplifying whole mitochondrial genomes for *A. a. azarai* and *A. nancymaae.*

		Aotus azarai azarai		*Aotus nancymaae*	
Primer Name	**Location ◊**	**Annealing (°C)**	**Extension**	**Annealing (°C)**	**Extension**
01-F	167	49Δ14	1 min, 45 sec	52	2 min
01-R	899	49Δ14	1 min, 45 sec	52	2 min
02-F	709	49Δ14	1 min, 45 sec	52	2 min
02-R	1576	49Δ14	1 min, 45 sec	52	2 min
02-Fa	682	51.1 – 57.3	1 min, 30 sec	52	2 min
02-Ra	1432	51.1 – 57.3	1 min, 30 sec	52	2 min
03-F	1376	49Δ14	1 min, 45 sec	52	2 min
03-R	2136	49Δ14	1 min, 45 sec	52	2 min
04-F	2015	50.0 – 57.3	1 min, 45 sec	52	2 min
04-R	2938	50.0 – 57.3	1 min, 45 sec	52	2 min
05-F	2819	42.3 – 57.3	1 min, 45 sec	52	2 min
05-R	3663	42.3 – 57.3	1 min, 45 sec	52	2 min
06-F	3560	50Δ14	1 min, 45 sec	52	2 min
06-R	4951	50Δ14	1 min, 45 sec	52	2 min
07-F*	4082	42.3 – 54.2	1 min, 45 sec	42.6 – 57.4	1 min, 30 sec
07-R*	5142	42.3 – 54.2	1 min, 45 sec	42.6 – 57.4	1 min, 30 sec
07-Fa	4044	50Δ20	2 min, 30 sec	52	2 min
07-Ra	5280	50Δ20	2 min, 30 sec	52	2 min
08-F	5031	50Δ14	1 min, 45 sec	52	2 min
08-R	6240	50Δ14	1 min, 45 sec	52	2 min
09-F	5626	50Δ14	1 min, 45 sec	50Δ20	1 min, 30 sec
09-R	7000	50Δ14	1 min, 45 sec	50Δ20	1 min, 30 sec
COII-F2	6954	50Δ20	1 min	52	2 min
COII-R2	7722	50Δ20	1 min	52	2 min
10-F	7433	50Δ14	1 min, 45 sec	52	2 min
10-R	8314	50Δ14	1 min, 45 sec	52	2 min
11-F	8245	50Δ16	2 min	52	2 min
11-R	8894	50Δ16	2 min	52	2 min
12-F	8783	50Δ16	2 min	52	2 min
12-R	9545	50Δ16	2 min	52	2 min
13-F*	9359	50Δ16	2 min	52	2 min
13-R*	10523	50Δ16	2 min	52	2 min
13-Fa	9166	46.2-51.4 without DMSO 51.4-54.1 with DMSO∇	2 min, 30 sec; 40 cycles	52	2 min
13-Ra	10475	46.2-51.4 without DMSO 51.4-54.1 with DMSO∇	2 min, 30 sec; 40 cycles	52	2 min
14-F*	9545	50Δ16	2 min	52	2 min
14-R*	11398	50Δ16	2 min	52	2 min
14-Fa*	9566	54.1-57.4, without DMSO 57.4 with DMSO∇	2 min, 30 sec; 40 cycles	52	2 min
14-Ra*	11427	54.1-57.4, without DMSO 57.4 with DMSO∇	2 min, 30 sec; 40 cycles	52	2 min
14-Fb	10349	51.4 – 59.2 without DMSO 42.6 – 59.2 with DMSO∇	1 min	52	2 min
14-Rb	11239	51.4 – 59.2 without DMSO 42.6 – 59.2 with DMSO∇	1 min	52	2 min
15-F	11154	50Δ16	2 min	52	2 min
15-R	12434	50Δ16	2 min	52	2 min
16-F	12280	47.0 – 57.3	2 min	F	F
16-R	13606	47.0 – 57.3	2 min	F	F
16-Fa	12354	N/A	N/A	48.9 – 54.1	1 min, 30 sec; 35 cycles
16-Ra	13592	N/A	N/A	48.9 – 54.1	1 min, 30 sec; 35 cycles
17-F	13338	47.0 – 57.3	2 min	F	F
17-R	14308	47.0 – 57.3	2 min	F	F

Primer Name	Location ◊	*Aotus azarai azarai* Annealing (°C)	*Aotus azarai azarai* Extension	*Aotus nancymaae* Annealing (°C)	*Aotus nancymaae* Extension
17-Fa	13414	N/A	N/A	50Δ20	1 min, 30 sec; 35 cycles
17-Ra	14326	N/A	N/A	50Δ20	1 min, 30 sec; 35 cycles
18-F	14066	51.1 – 57.3	2 min	52	2 min
18-R	15418	51.1 – 57.3	2 min	52	2 min
19-F	14549	50Δ16	2 min	†	†
19-R	15823	50Δ16	2 min	†	†
19-Fa	14636	N/A	N/A	†	†
19-Ra	15839	N/A	N/A	†	†
20-F	16072	50Δ16	2 min	52	2 min
20-R	492	50Δ16	2 min	52	2 min
CR1-F	15418	50Δ20	1 min	•	•
CR1-R	15823	50Δ20	1 min	•	•
CR2-F	15823	50Δ20	1 min	•	•
CR2-R	16072	50Δ20	1 min	•	•
CR3-F	16067	50Δ20	1 min	52	2 min
CR3-R	16486	50Δ20	1 min	52	2 min

Components for all PCR reactions (unless otherwise noted): 13.25 µL ddH_20, 2.5 µL, Buffer II (ABI), 2 µL $MgCl_2$ (ABI), 1 µL 10 mM dNTPs pre-mix (Invitrogen), 1 µL forward primer, 1 µL reverse primer, 0.25 µL AmpliTaq Gold (ABI), 4 µL DNA (5 ng/µL).

◊Denotes the position of the 5'-base relative to the published mitochondrial sequence for *A. lemurinus* (Hodgson et al., 2009).

*Primers that failed to work during amplification PCRs are denoted by an asterisk. Such primers were re-designed, and the re-designed primers are denoted by the suffix –a or –b at the end of the primer name. Note that for Region 14, primer pair (14-Fa/14-Ra) also failed in the amplification PCR; hence the pair with the –b suffix (14-Fb/14-Rb) represents the pair actually used in amplification and sequencing for Region 14 of both mitochondrial genomes.

∇Dimethyl sulfoxide (DMSO) was spiked in at 5% of the total reaction volume (1.25 µL per 25 µL reaction) to facilitate correct priming and amplification of the specified regions. The amount of ddH_2O was adjusted accordingly.

"N/A" indicates the primers were not tested against that species.

"F" indicates the primers failed against that species.

•The control region for *A. nancymaae* was amplified according to the parameters set forth above in **Table 2.3**.

† Specific parameters for amplifying mtDNA region 19 in *A. nancymaae* were as follows: initial denaturation at 95 °C for 5 minutes, 38 cycles of (heating at 95 °C for 30 seconds; annealing at 52.5 °C for 45 seconds; extension at 72 °C for 2 minutes 10 seconds), final extension at 72 °C for 7 minutes, and final hold at 4 °C. Amplification of region 19 was achieved with primers 19-F/CR3-R and 19-Fa/CR3-R. Sequencing reactions used the primers 19-Fa, CR2-F (forwards) and 19-R, 19-Ra, CR2-R, CR3-R (reverses), all with standard sequencing parameters.

Amplicons were purified by SAP/Exo I digestion (New England BioLabs) and cycle-sequenced using Big Dye™ v3.1 (ABI). Excess dye terminators were removed with the BigDye XTerminator™ purification kit (ABI), and DNA sequences were read on a 3130xl Gene Analyzer (ABI). Read quality of chromatograms was assessed using Sequencing Analysis v5.4 software (ABI), and bidirectional sequences were aligned and assembled using Sequencher v4.9 (Gene Codes) and Geneious Pro v5.16 (Drummond et al., 2010).

All new mtDNA sequences from this study have been deposited in GenBank, and their accession numbers are listed in **Table 2.4**.

Table 2.4. CR and COII haplotype definitions for all individuals and species and subspecies analyzed.

Taxon	Common Name	CR Hg	COII Hg	Freq.	Animal location	GenBank
Aotus azarai azarai	Azara's Owl Monkey	A*	AaaØ*	26	Wild	JN161069, JN161062
		A*	AaaI*	1	Wild	JN161069, JN161063
		A1*	AaaØ*	1	Wild	JN161070, JN161062
		A1a*	AaaØ*	1	Wild	JN161071, JN161062
		A2*	AaaØ*	2	Wild	JN161072, JN161062
		A4*	AaaØ*	1	Wild	JN161073, JN161062
		A5*	AaaØ*	1	Wild	JN161074, JN161062
		A6*	AaaØ*	1	Wild	JN161075, JN161062
		B*	AaaØ*	32	Wild	JN161076, JN161062
		B1*	AaaØ*	1	Wild	JN161077, JN161062
		B1a*	AaaØ*	1	Wild	JN161078, JN161062
		B2*	AaaØ*	1	Wild	JN161079, JN161062
		B3*	AaaØ*	1	Wild	JN161080, JN161062
		B4*	AaaØ*	1	Wild	JN161081, JN161062
		B5*	AaaØ*	1	Wild	JN161082, JN161062
		B6*	AaaØ*	1	Wild	JN161083, JN161062
		B7*	AaaØ*	1	Wild	JN161084, JN161062
		B8*	AaaØ*	1	Wild	JN161085, JN161062
		C*	AaaØ*	8	Wild	JN161086, JN161062
		C*	AaaII*	1	Wild	JN161086, JN161064
		C1*	AaaØ*	13	Wild	JN161087, JN161062
		C1*	AaaIII*	3	Wild	JN161087, JN161065
		C1a*	AaaIII*	1	Wild	JN161088, JN161065
		C2*	AaaØ*	7	Wild	JN161089, JN161062
		C2*	AaaIV*	1	Wild	JN161089, JN161066
		C2a*	AaaØ*	1	Wild	JN161090, JN161062
		C2c*	AaaØ*	1	Wild	JN161091, JN161062
		C3a*	AaaØ*	1	Wild	JN161092, JN161062
		C3b*	AaaV*	1	Wild	JN161093, JN161067
		C4*	AaaØ*	1	Wild	JN161094, JN161062
		C5*	AaaØ*	1	Wild	JN161095, JN161062
		X*	AaaVI*	1	Wild	JN161096, JN161068
		Y*	AaaØ*	1	Sáenz Peña Zoo	JN161097, JN161062

Taxon	Common Name	CR Hg	COII Hg	Freq.	Animal location	GenBank
		Z*	AaaØ*	1	Sáenz Peña Zoo	JN161098, JN161062
Aotus azarai boliviensis	Bolivian Owl Monkey	-	Aab01 [a]	1	GenBank	U36846
Aotus infulatus	Feline Owl Monkey	-	Aai04 [b]	1	GenBank	DQ321662
		-	Aai05 [b]	1	GenBank	DQ321663
		-	Aai09 [b]	1	GenBank	DQ321667
		-	Aai10 [b]	1	GenBank	DQ321668
Aotus lemurinus	Grey-bellied Owl Monkey	-	Al001*	1	San Diego Zoo	JN161047
		-	Al01 [a]	1	GenBank	U36845
		-	Al02 [a]	1	GenBank	U36844
		-	Al03 [a]	1	GenBank	U36843
Aotus lemurinus brumbacki	Brumback's Owl Monkey	-	Alb11 [b]	1	GenBank	DQ321669
Aotus lemurinus grisiemembra	Grey-handed Owl Monkey	-	Alg002*	1	San Diego Zoo	JN161048
		-	Alg03 [b]	1	GenBank	DQ321661
Aotus nancymaae	Nancy Ma's Owl Monkey	-	Ana001*	1	San Diego Zoo	JN161049
		-	Ana01 [b]	1	GenBank	DQ321659
		-	Ana02 [b]	1	GenBank	DQ321660
		-	Ana12113 [c]	1	GenBank	AF352255
		-	Ana837 [c]	1	GenBank	AF352254
		-	Ana04 [a]	1	GenBank	U36770
		-	Ana002*	1	DuMond	JN161050
		-	Ana003*	1	DuMond	JN161051
		-	Ana004*	1	DuMond	JN161052
		-	Ana005*	1	DuMond	JN161053
		-	Ana006*	1	DuMond	JN161054
Aotus nigriceps	Black-headed Owl Monkey	-	Ani001*	1	San Diego Zoo	JN161055
		-	Ani12012 [c]	1	GenBank	AF352258
		-	Ani718 [c]	1	GenBank	AF352256
		-	Ani719 [c]	1	GenBank	AF352257
Aotus trivirgatus	Three-striped Owl Monkey	-	At01 [d]	1	GenBank	AY250707
Aotus vociferans	Spix's or Noisy Owl Monkey	-	Av07 [b]	1	GenBank	DQ321665
		-	Av08 [b]	1	GenBank	DQ321666
		-	Av328 [c]	1	GenBank	AF352259
		-	Av331 [c]	1	GenBank	AF352260
Callicebus donacophilus	White-eared Titi Monkey	-	Cd001*	1	San Diego Zoo	JN161056
		-	Cd002*	1	San Diego Zoo	JN161057
Pithecia pithecia pithecia	White-faced Saki Monkey	-	Ppp001*	1	San Diego Zoo	JN161058
		-	Ppp002*	1	San Diego Zoo	JN161059
Saimiri sciureus sciureus	Common Squirrel Monkey	-	Sss001*	1	San Diego Zoo	JN161060
		-	Sss002*	1	San Diego Zoo	JN161061
Tarsius syrichta	Philippine Tarsier	-	Ts001 [e]	1	GenBank	L22784
Macaca mulatta	Rhesus Macaque	-	Mm001 [f]	1	GenBank	M74005
Lemur catta	Ring-tailed Lemur	-	Lc001 [e]	1	GenBank	AJ421451

Locations: Wild animals sampled in Formosa, Argentina (Lat = 25°, 59.4' South; Long = 58°, 11.0' West).
Data Sources: *current study, [a] Ashley and Vaughn, 1995, [b] Plautz et al., 2009, [c] Suarez et al., unpublished 2001, [d] Collura et al., unpublished 2003, [e] Arnason et al., 2002, [f] Disotell et al., 1992

Phylogenetic analyses

To explore the phylogeographic origin of *A. a. azarai* and its relationship to other owl monkey species, we conducted phylogenetic analyses of COII sequences. Searches in GenBank (NCBI) yielded 23 informative COII sequences representing nine of the fourteen putative species and subspecies usually described when discussing the genus *Aotus* (**Table 2.1**). COII sequences for *A. miconax*, *A. hershkovitzi,* and *A. zonalis* were

unavailable. Because several of the GenBank sequences lacked the entire 696 bp gene sequence, the COII sequence matrix was pruned to 549 bases (nt positions 16-564) for interspecies analyses. Final COII alignments were also translated (vertebrate mtDNA code) to check for stop codons, the presence of which could indicate the amplification of numts rather than true mtDNA sequences.

The 118 *A. a. azarai* sequences were restricted to the seven unique haplotypes identified in the study population to minimize phylogenetic errors (Zwickl and Hillis, 2002), and then combined with 23 COII sequences from other species of *Aotus* obtained from GenBank, along with the nine non-*azarai* owl monkey sequences generated for this study. The set of 39 *Aotus* COII sequences represented 12 owl monkeys from taxa distributed north of the Amazon River, 11 from the west, and 16 from the south. The six sequences generated from *Callicebus*, *Pithecia*, and *Saimiri* samples were included to obtain a greater range of coalescent points within the platyrrhines, and sequences from single *Macaca*, *Tarsius*, and *Lemur* individuals were used as outgroups.

To select the most appropriate model for our phylogenetic analyses, we ran the program jModelTest v0.1.1 (Guindon and Gascuel, 2003; Felsenstein, 2005; Posada, 2008) using 11 substitutions patterns to survey 88 models of nucleotide substitution (+F base frequencies, rate variation of +I and +G with nCat = 4). The modified/corrected Akaike Information Criterion (AICc) setting was implemented because of the small size of comparative nucleotide characters (549), and parallel searches using Bayesian Information Criterion (BIC) and performance-based decision theory (DT) were conducted. The base tree for our likelihood calculations was optimized for Maximum Likelihood (ML) phylogenetic analysis. All three searches in jModeltest selected the

TPMuf1+G model (Kimura, 1981) with a likelihood score (-lnL) of 3370.84. This model was applied to the Maximum Likelihood (ML) analysis implemented in the phylogenetic program PAUP* 4.0b10 (Swofford, 2002) to maximize the probability of observing the alignment of *Aotus* COII nucleotides. Bootstrap values were estimated based on a set of 10,000 replicates.

Bayesian inference analysis (BI) was undertaken to obtain the most probable set of trees given an evolutionary model and our specific alignment of data, using the software program BEAST v1.5.3 (Drummond and Rambaut, 2007). In addition to BI phylogenetic tree construction, BEAST can estimate coalescent dates using a relaxed lognormal molecular clock that accounts for post-divergence and lineage-specific variations in mutation rate (Drummond et al., 2002, 2006; Ho et al., 2005).

We imported 48 COII sequences (34 *Aotus*, 2 *Callicebus*, 2 *Pithecia*, 2 *Saimiri*, 1 *Macaca,* 1 *Tarsius*, and 1 *Lemur*) into the program BEAUti v1.5.3 to format the run file for BEAST. This set of taxa was used to provide an adequate evolutionary time depth that encompassed the few known fossil platyrrhines (**Table 2.5**). As the TPMuf1+G model selected by jModelTest was unavailable in BEAST, we used the TN93 substitution model with three partitions for codon positions, empirical base frequencies. We implemented a randomly generated starting tree and the Yule Process speciation parameter as the tree prior. We specified the fossil time points as log-normally distributed priors applied to the appropriate taxon designations (**Figure 2.3**). The Markov Chain Monte Carlo (MCMC) search was run with four chains for 10,000,000 generations, with trees sampled every 1000 generations.

Table 2.5. Fossil calibration points used to estimate time to most recent common ancestor (TMRCA)

Clade	Fossil	Epoch	Radiometric Dates (mya)	Shape	Calibration Mean (mya)	SD (mya)	Offset (mya)
Aotus	*Aotus dindensis*[a]	Miocene	11.8-13.5	Lognormal	-1.0	0.85	6.5
Saimiri	*Neosaimiri*[b]	Miocene	12-15	Lognormal	1.0	1.0	1.0
Platyrrhini	*Branisella bolivana*[c]	Miocene	27	Lognormal	1.0	0.5	22
Platyrrhini + Catarrhini	*Parapithecus grangeri*[d]	Oligocene	36-40	Lognormal	1.0	0.5	30
	Catopithecus browni[e]	Eocene	30-36				
	Proteopithecus sylviae[f]	Eocene	36				
Primates	*Plesiadapiforms*[g]	Paleo/Eocene	62	Lognormal	1.0	0.5	59
	K-T extinction event[g]	Paleocene	65				

Data Sources: [a] Setoguchi and Rosenberger, 1987, [b] Hartwig and Meldrum, 2002, [c] Takai et al., 2000, [d] Beard and Wang, 2004, [e] Simons et al. 1987, [f] Takai and Ayana, 1996, [g] Bloch et al., 2007.

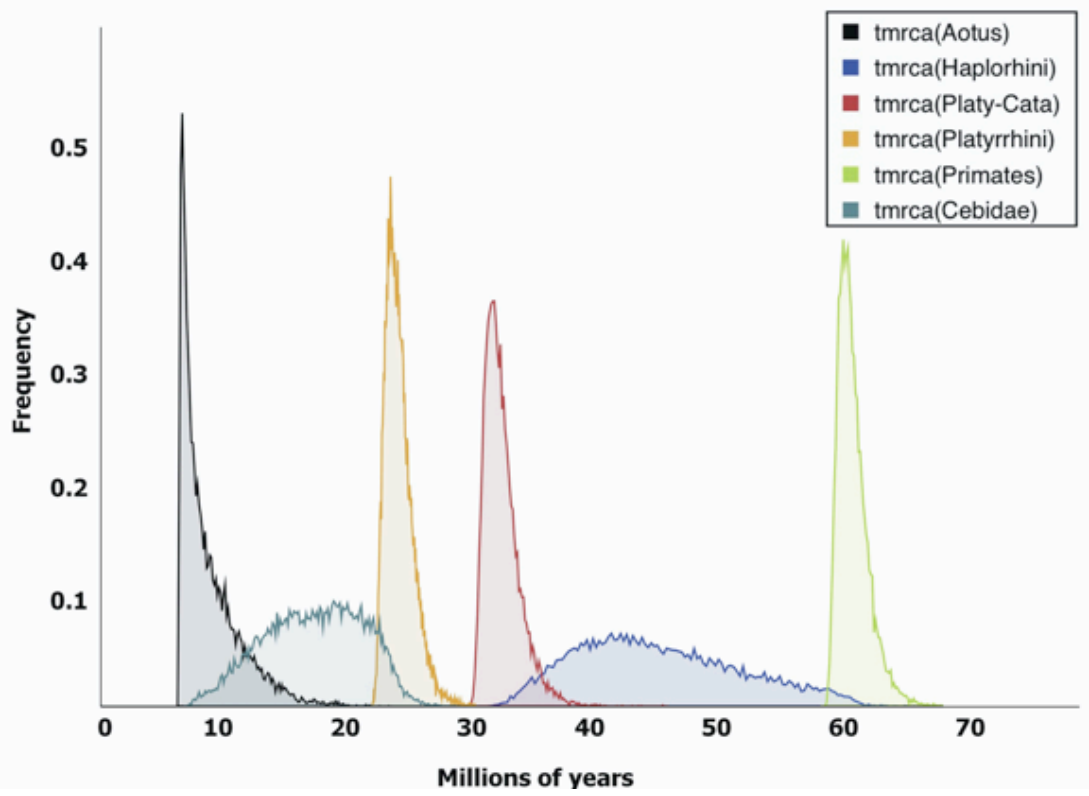


Figure 2.3. Marginal density plots of TMRCA age estimates for different taxon-based prior constraints utilized in the interspecific Bayesian age estimations. The frequency distributions of age ranges for each taxonomic grouping were calculated through 10,000,000 generations of the TPM1uf1+I+G model (Kimura, 1981) in BEAST v1.5.3, using priors described in **Table 2.5**.

Using the average standard deviation in split frequencies among the four chains (0.01), we assessed the level of convergence (<0.05) to represent an acceptable level of post-convergence tree likelihoods, which influence the accuracy of our consensus Bayesian tree. The first 1000 trees were discarded as "burn-in" to remove extraneous pre-convergence probability values (Altekar et al., 2004). We further analyzed the results generated in BEAST in TRACER v.1.5 (Rambaut and Drummond, 2007) to assess the accuracy of the estimations based on the effective sample sizes (ESS) of our data.

Output files from the ML and BI phylogenetic calculations were summarized using TreeAnnotator v1.5.3 (Drummond and Rambaut, 2007) to construct a consensus tree (50%) on the basis of mean node heights and maximum clade credibility values. Summary trees were imported into FigTree v1.2.3 (Drummond and Rambaut, 2007) for visualization. Using MacClade v4.0.8 (Madison and Madison, 2003), the consensus (CI), rescaled consistency (RC) and retention indices (RI) were also assessed.

Population genetic analyses

To assess species and population-level mtDNA variation, summary statistics, including gene (π) and haplotype (h) diversity, were calculated for both the COII and CR sequences using programs available in Arlequin v3.11 (Excoffier et al., 2005) and DnaSP v4.50 (Rozas et al., 2003). Both data sets were also used to conduct pairwise mismatch analysis, and calculate expansion variables (τ), which provide estimates of past population size and dynamics (Rogers and Harpending, 1992). In addition, the neutrality indices Tajima's *D* (Tajima, 1989a, b) and Fu's F_S (Fu, 1997) were calculated to estimate whether population expansions or contractions had occurred. Transversions (TV) were

weighted higher than transitions (TI) (10:1) in all calculations to account for the differential probability of either occurring across evolutionary time (Excoffier et al., 2005).

Multistate median joining (MJ) networks were generated from both COII and CR sequences using NETWORK v4.5.02 (Bandelt et al., 1999) to investigate the intraspecific phylogenetic relationships among samples based on parsimony (Posada and Crandall, 2001). In the construction of networks with the CR data set, a 10:1 TV:TI weighting scheme derived from human control region studies was employed (Bandelt and Parson, 2008). This ratio does not deviate significantly from the 9.5:1 TV:TI ratio previously estimated in primate mtDNA studies (Yoder et al., 1996; Purvis and Bromham, 1997; Yang and Yoder, 1999). In addition to this weighting scheme, the characters at CR nucleotide positions 136, 210, 249, 256, 275, 851 and 927 were down-weighted to reduce the reticulations caused by these hypervariable characters. The caveat inherent to working with character-based networks is that no alternative evolutionary model other than parsimony can be tested.

To explore the demographic history of the study population, pairwise differences among the *A. a. azarai* sequences in both the CR and COII data sets were calculated, and the frequency distributions of observed pairwise mismatches were plotted. For each of the analyses outlined above, the mean number of pairwise differences and raggedness index were calculated (Rogers and Harpending, 1992).

For intraspecific phylogenetic dating, the rho value (ρ), a product of mutation rate (μ) and time (τ), was determined. The rho value reflects the average number of pairwise differences between a set of sequences to a designated root. This value was estimated for

all distinct clusters and their sub-branches in the CR data set using NETWORK v4.5.02 (Bandelt et al., 1995, 1999).

The within-population rate of mutation for the *Aotus* mitochondrial CR remains uncertain. Thus, we utilized two different CR mutation rates (ω: changes per site per million years) to provide high and low estimates of intraspecific genetic coalescence for the study owl monkey population. These two rates were drawn from human (ω: 0.320, Sigurgardóttir et al., 2000) and primate (ω: 0.111, Weinreich, 2000) studies. The pedigree-based human mutation rate translates into 2,843.49 years per CR mutation, whereas the population-based primate mutation rate translates into 8,197.39 years per CR mutation. It should be noted that the 95% confidence intervals and standard errors for coalescent dates do not consider mutation rate errors (Forster et al., 1996).

2.3 Results

Whole mtDNA genome sequencing of A. a. azarai and A. nancymaae

We assessed molecular variation at the genus level by examining whole mitochondrial genome sequences from two previously unexamined *Aotus* species, Azara's and Nancy Ma's owl monkeys (*A. a. azarai*: JN161099; *A. nancymaae*: JN161100). When compared with the mitochondrial genome of *A. lemurinus* (FJ785421, Hodgson et al., 2009) as a reference sequence, we observed a large number of single nucleotide polymorphisms (SNPs) among all three species. *A. a. azarai* and *A. lemurinus* were distinguished from one another by 941 SNPs, *A. a. azarai* from *A. nancymaae* by 985 SNPs, and *A. lemurinus* from *A. nancymaae* by 835 SNPs.

In addition, we observed a striking difference in the sequence composition of the

CR of *A. nancymaae* relative to those of *A. a. azarai* and *A. lemurinus* (16,472 bp vs 16,585 bp in *A. a. azarai* and 16,580 bp in *A. lemurinus*). In *A. nancymaae*, ~13 separate deletions, ranging in size from 2 - 32 bp, shortened the genome by 113 bp, making it the smallest platyrrhine mitochondrial genome present in GenBank (including sequences from Hodgson et al., 2009). These deletions were observed in multiple *A. nancymaae* individuals through direct sequencing, and confirmed through PCR amplification and gel electrophoresis, with CR amplicons from *A. nancymaae* individuals being appreciably smaller than those of other owl monkey taxa (**Figure 2.4**).

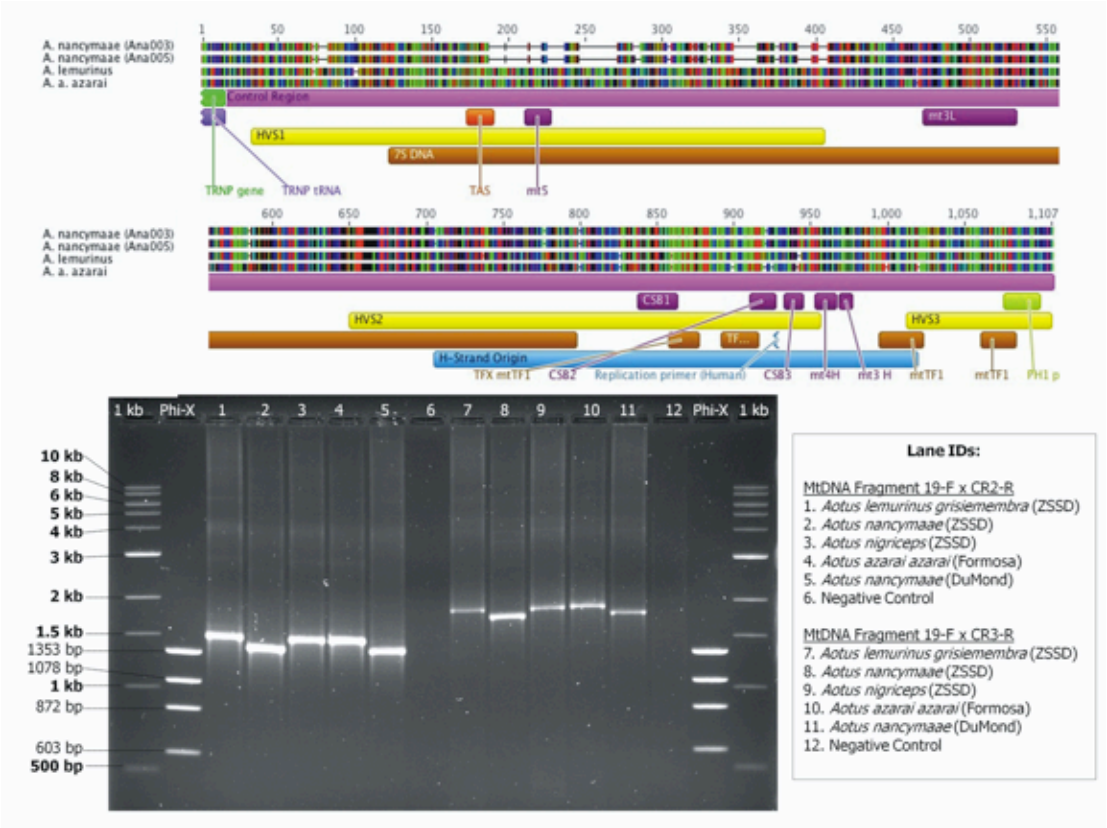


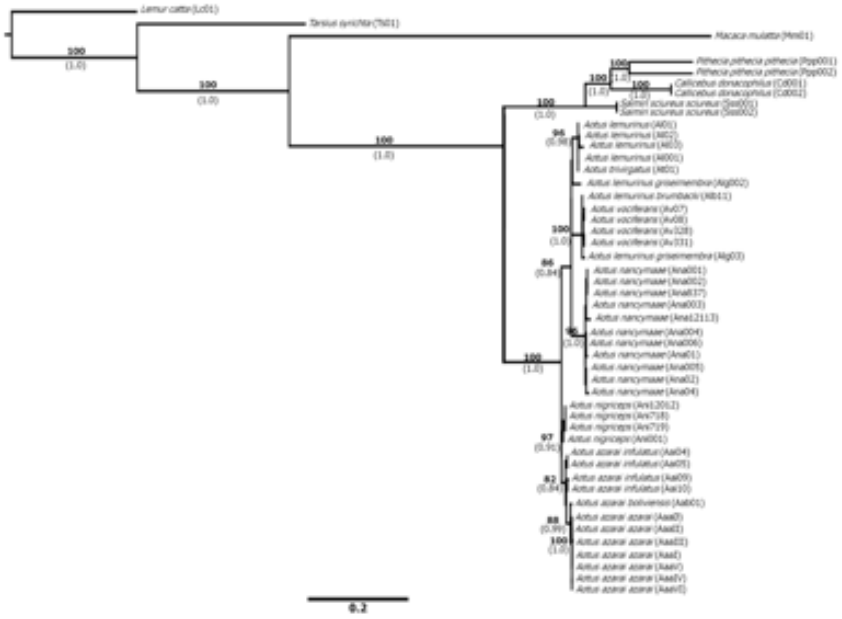
Figure 2.4. Control region deletions identified in the *A. nancymaae* mitochondrial genome, relative to *A. a. azarai* and *A. lemurinus*. The top portion of the figure diagrams the occurrence of deletions in *A. nancymaae* compared to other *Aotus spp.* at the nucleotide level, as observed through direct sequencing and subsequent comparative alignment. The bottom portion of the figure displays a gel electrophoresis photo depicting the smaller fragment lengths of *A. nancymaae* resulting from their CR deletions, as generated by two different primer pairs.

Phylogenetic analyses

We noted similar phylogenetic relationships within and among the northern, western, and southern clades of *Aotus* species irrespective of the ML or BI methods employed. The COII phylogeny showed a deep phylogenetic split between *Aotus* species and subspecies living north of the Amazon River and those living south of it (**Figure 2.5a**). The ML bootstrap values associated with this split were high (86 for the first clade of the bifurcation, 97 for the second), and the BI analysis exhibited similarly high posterior probabilities (1.00) for the same clusters of closely related species of *Aotus*.

We estimated the time to the most recent ancestor (TMRCA) at each phylogenetic node (**Figure 2.5b**). The TMRCA estimates varied widely, from 1.78 mya for *A. a. azarai* (95% HPD: 0.24 – 3.99 mya) to 4.68 mya for *A. nancymaae* (95% HPD: 1.93 – 8.10 mya) (**Table 2.6**). When dated according to their geographic ranges relative to the Amazon River, the TMRCA of northern species was 7.34 mya, whereas that of the different southern species was 6.22 mya. The TMRCA for the genus *Aotus* was 8.95 mya.

Panel a



Panel b

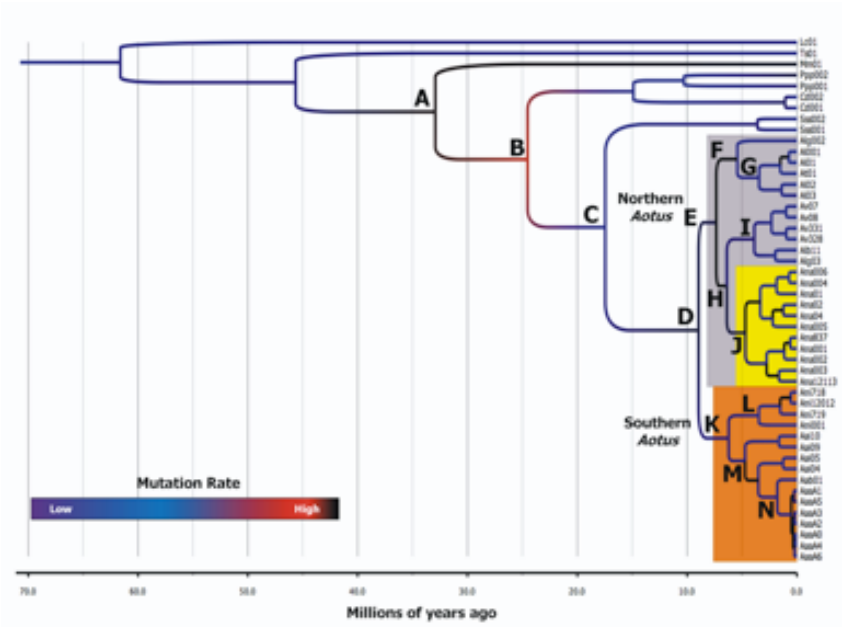


Figure 2.5, Panel a. A consensus phylogram of COII sequences from the genus *Aotus* constructed using two different tree-building methodologies. Branching patterns and branch lengths are based on the consensus of ML and BI phylogenetic analyses. ML bootstrap values >70% from 10,000 replicates are shown above the branches. Bayesian posterior probabilities are listed below branches in parentheses (BI). The ML phylogram is the consensus of 2,990 trees, with a length of 344 steps, CI = 0.66, RI = 0.82, RC = 0.54. The BI phylogram was based upon four computational chains run in parallel across 10,000,000 generations, with trees sampled every 1000 generations. **Panel b.** Bayesian chronogram depicting the coalescence times for 48 primate COII sequences. The mutation rate of COII is displayed in a blue (0 = low) to red (0.5 = high) color gradient along branches (color version available online). Arabic letters located at nodes refer to the splits at which coalescence dates were estimated in **Table 2.6**.

Table 2.6. Molecular dating of COII phylogenetic branches in the genus *Aotus*.

Node#	Age^	Confidence Interval*	Notable Events
A	32,936,700	30,733,500 – 35,779,000	Platyrrhine – catarrhine divergence
B	24,528,000	22,704,700 – 26,750,200	Coalescence of platyrrhines
C	17,452,700	9,904,300 – 24,412,800	Coalescence of Cebidae, emergence of *Aotus*
D	8,953,100	6,526,500 – 13,879,400	Coalescence of all extant taxa from genus *Aotus*
E	7,340,500	4,385,800 – 11,716,700	Coalescence of all "Northern" *Aotus* species
F	5,432,000	**	Divergence of *A. lemurinus* and *A. l. griseimembra*
G	3,440,900	600,900 – 6,881,900	Coalescence of *A. lemurinus*
H	6,382,100	**	Divergence of *A. vociferans* and *A. nancymaae*
I	2,307,500	410,000 – 4,799,400	Coalescence of *A. vociferans*
J	4,682,300	1,929,100 – 8,103,700	Coalescence of *A. nancymaae* ("Western migration")
K	6,215,600	2,777,100 – 10,516,200	Coalescence of all "Southern" *Aotus* species
L	3,461,800	466,300 – 7,360,500	Coalescence of *A. nigriceps*
M	4,711,900	1,482,800 – 8,464,800	Derivation of *A. infulatus* population(s)
N	1,782,200	235,400 – 3,989,100	Derivation of *A. a. boliviensis*, coalescence of *A. a. azarai*

#Letter designations refer to node identifications on the Bayesian chronogram in **Figure 2.5b**.
^All COII dates are rounded to the nearest 100.
*95% HPD confidence intervals listed as: lower bound - upper bound. See **Figure 2.3** for marginal density plots of TMRCA range distributions.
** Represents a split clade. For split clades and single-frequency taxa, confidence intervals could not be estimated.

Sequence diversity and population structure in A. a. azarai

The mtDNA CR sequence analysis revealed considerable genetic diversity in the Azara's owl monkey study population. Fifty-two polymorphic sites (TI, TV, and insertion-deletion substitutions) were present in 118 individuals, and they defined 30 distinct haplotypes (<5% missing data; **Table 2.7**). We also obtained full COII sequences for all 118 individuals, and observed that eight polymorphic nucleotides defined seven unique COII haplotypes. The TV:TI ratio was 4:4 for the eight segregating sites, but transitions were more frequent at the third nucleotide position.

DNA sequences for the COII gene were well conserved when compared to those of the CR (**Table 2.7**). The Nei's gene diversity (π) was 0.005 for CR haplotypes, but only 0.001 for the COII haplotypes. Similarly, haplotype diversity estimates yielded

values of 0.83 for the CR sequences, and 0.14 for the COII sequences. However, both the CR and COII data sets had modest population expansion values (τ of 5.2 and 3, respectively) and negative Tajima's *D* and Fu's F_S values.

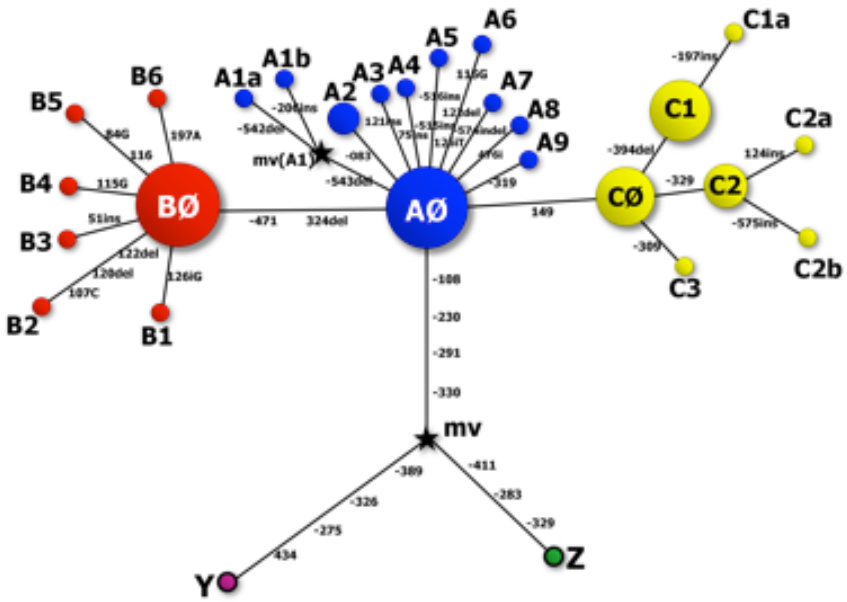
Table 2.7. Statistical indices of molecular diversity in *Aotus azarai azarai.*

	CR	COII
Summary Statistics		
Sample Size	118	118
Nucleotides (bp)	1099	696
Polymorphic Sites	52	8
Transitions (TI)	23	4
Transversions (TV)	5	4
Insertion/Deletions	25	0
Haplotypic Diversity		
# Haplotypes	30	7
Haplotype Diversity (*h*)	0.83	0.14
Sequence Diversity		
Nei's Gene Diversity (π)	0.005 ± 0.002	0.001 ± 0.001
Tau (τ)	5.2	3
Mean # Pairwise Differences	5.6 ± 2.7	0.8 ± 0.6
Harpending's Raggedness Index	0.01	0.6
p (Harpending's)	0.7	0.6
Selective Neutrality		
Tajima's *D* (1000 simul.)	-0.4	-1.1
p (*D* simul < *D* obs)	ns	0.001
Fu's F_S (1000 simul.)	-25.2	-3.4
p (sim. $F_S \leq$ obs. F_S)	ns	ns

The network of CR sequences contained three distinct clades, or haplogroups (hg) (**Figure 2.6a**). Each clade consisted of a central, high frequency haplotype and a number of derivative haplotypes extending from it. The CR network also included three outlier haplotypes that correspond to three animals that appeared to be distantly related to the

other individuals within the population. They included a solitary individual captured within the study area (X) and two zoo animals (Y and Z). The intraspecific network generated from the seven unique COII haplotypes was less structured than the one based on CR sequences (**Figure 2.6b**). COII haplotype "AaaØ" represented 93% of the individuals (109 of 118), and the remaining haplotypes were only 1-2 mutational steps away from this founder type.

Panel a



Panel b

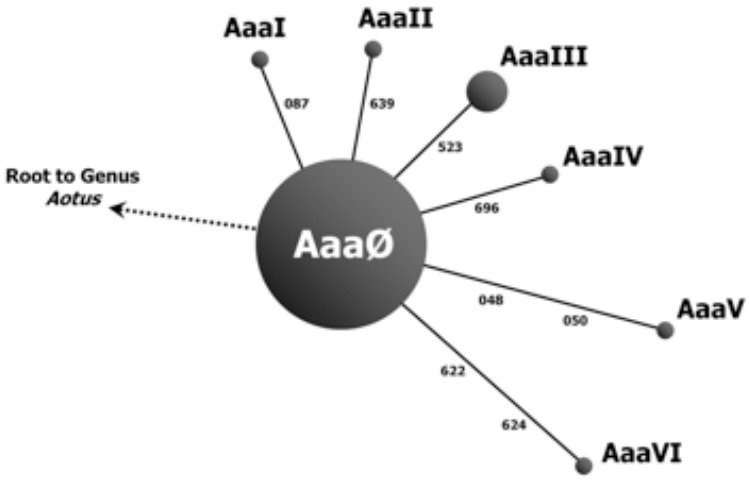


Figure 2.6, Panel a. Median joining network of *Aotus azarai azarai* CR sequences. The three major clades of CR sequences are designed A, B, and C. Founder haplotypes are given the suffix "Ø," and derivative haplotypes are numbered sequentially from them. The haplotype associated with each *A. a. azarai* sample is indicated in Table 2. Haplotype X represents a solitary individual captured in the study area, while Y and Z represent individuals from a local zoo. Putative intermediate haplotypes are designated as median vectors (mv) in the network. The specific mutations (homologous to human mtDNA positions 579 bases before (denoted [-]) and 509 bases after the "0" point or origin of replication) that define the haplotypes are shown along the branches of the network. **Panel b.** Median joining network of *A. a. azarai* COII sequences. The founder haplotype is given the suffix "Ø" and derivative haplotypes are numbered sequentially from it using Roman numerals to distinguish this network from the relationships defined by the CR data.

The mismatch analyses provided details about the demography of southern Azara's owl monkeys. The mismatch distribution for CR sequences of *A. a. azarai* showed a relatively small number of pairwise differences, with the curve being strongly skewed to the left (**Figure 2.7**). The COII sequences for the same individuals displayed a similar left-skewed mismatch curve, albeit at lower resolution. This limited diversity of COII sequences was reflected in the raggedness index (**Table 2.7**), as well as the COII haplotype network (**Figure 2.6b**).

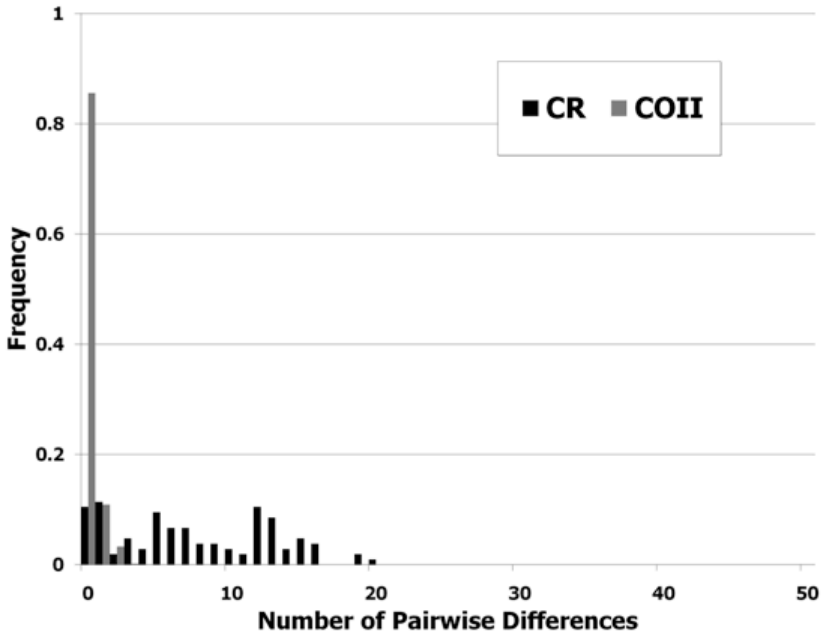


Figure 2.7. Pairwise mismatch distributions of *A. a. azarai* CR and COII sequences.

We estimated dates for the primary maternal lineages appearing in the study population using information from the network analyses (**Table 2.8**). Of the three major CR haplogroups, hg-A appeared to be the oldest, due to its central location in the network and the presence of two derived haplogroups (hg-B and hg-C) extending from it. We confirmed this impression through the calculation of ρ for each haplogroup relative to its

putative ancestral node using two different rates for ω (human: 0.320 vs. primate: 0.111). For hg-B and hg-C, the ancestral node was hg-A, whereas for hg-A, the ancestral node was a median vector (mv) that linked it to the more divergent haplotypes Y and Z that belonged to two zoo animals. The lack of a transitional haplotype directly ancestral to hg-A may have inflated the coalescence estimate for that clade. Even so, these estimates generated coalescence dates of 12,161 years before present for hg-A, 3,888 years before present for hg-B, and 3,303 years before present for hg-C when the faster, pedigree-based human ω is used. The CR haplogroup age estimations increase to 35,065 years before present for hg-A, 11,211 years before present for hg-B, and 9,523 years before present for hg-C when the slower, population-based primate ω is applied.

Table 2.8. Molecular dating of CR haplogroup (hg) clades in *Aotus azarai azarai.*

					Human rate (ω: 0.320) [a]			Primate rate (ω: 0.111) [b]		
Clade	**Ancestral node***	**Descendant nodes (*n*)****	**Age in Mutations (ρ)**	**SD (σ)**	**Age (years)**	**SD (years)**	**S_E (years)**	**Age (years)**	**SD (years)**	**S_E (years)**
hg-A	mv	A0 - A7	4.28	2.00	**12162**	5694	2013	**35065**	16416	5804
hg-B	A0	B0 - B8	1.37	0.85	**3888**	2429	810	**11211**	7003	2334
hg-C	A0	C0 - C5***	1.16	0.67	**3303**	1917	578	**9523**	5528	1667

All CR dates are rounded to the nearest 10.
*The ancestral taxon is the node from which all mutational steps are counted to reach all inclusive descendant nodes in a clade.
**Descendant taxa are nodes believed to have arisen from an ancestral node.
***mvC3 was included in this haplogroup for coalescence estimates.
SD = Standard Deviation
S_E = Standard Error

Data Sources: [a] Sigurgardóttir et al., 2000, [b] Weinreich, 2000

2.4 Discussion

Origins of Aotus

Through our analyses of the mitochondrial genome, we have characterized the genetic variation of *A. a. azarai* at the genus, species, and population levels. By sequencing the two complete mitochondrial genomes *A. a. azarai* and *A. nancymaae*, we were able to compare total mtDNA diversity with the previously published *A. lemurinus* mitochondrial genome. This comparison reinforced the tripartite distinction of northern, southern, and western owl monkey species, and revealed a surprising series of CR mutations that have occurred in *A. nancymaae* since its split, and subsequent isolation from *A. lemurinus* and other northern species. Many of the CR polymorphisms specific to *A. nancymaae* were large deletions (>5 bp), the sum of which reduces the CR by 113 bp (**Figure 2.4**). While the deletions do not appear to affect the nucleotide composition of known conserved sequence blocks (CSBs 1-3), 7S DNA loop overlap or mitochondrial transcription factor (mtTF) binding sites, one deletion is situated close to the mitochondrial replication termination site (TAS). Although the functional impact of the *A. nancymaae* deletions is unknown, large CR deletions of this kind have been reported among species of platypus (Gemmell et al., 1996) and subspecies of gorillas (Xu and Arnason, 1996). The functional impact of these particular CR deletions is not well understood.

Chronology of COII Aotus phylogeny

Recent surveys of putative *Aotus* taxa delineated the deep phylogenetic split between northern and southern types, and identified the separation of *A. nancymaae* as

the result of a western migratory scenario (Plautz et al., 2009; Menezes et al., 2010). However, these studies estimated the coalescence of all modern *Aotus* lineages at ~4.0 mya. This estimate permits only a limited amount of time for the emergence of the karyotypic, morphological, and immunological differences that are observed among *Aotus* taxa today. Likewise, a relatively recent north-south owl monkey split would imply an inordinately long period of evolutionary stasis if recent fossil and molecular estimations of 10-12 mya for the emergence of the genus were accurate (Setoguchi and Rosenberger, 1987; Hodgson et al., 2009; Rosenberger et al., 2009; Takai et al., 2009).

Our analyses point to an older and more complex evolution of the genus (**Figure 2.8**). Based on the analysis of 39 COII haplotypes from ten different owl monkey taxa, we estimate that the TMRCA for the genus *Aotus* is 8.95 mya, a date that is considerably older than the previous coalescent date estimates of 3.6-4.7 mya (Ashley and Vaughn, 1995; Plautz et al., 2009). This discrepancy may be due to the increased number of putative taxa that were sampled for this study, phylogenetic outgroup rooting, or the use of fossils as calibration time points rather than molecular dating using internal rooting and mutation rates derived from an autosomal locus. Repeated BEAST runs with different parameters (nucleotide substitution models, normalized calibration points) consistently produced time frames similar to the results presented here. In any event, the TMRCA for genetic loci will always predate the actual splitting of populations.

We also observe a deep phylogenetic split between *Aotus* species living north of the Amazon River from those living to the south that may have begun over 8 mya. Following this split, *A. nancymaae* diverged from other northern groups and became genetically isolated. These clades also correlate with differences in malarial resistance

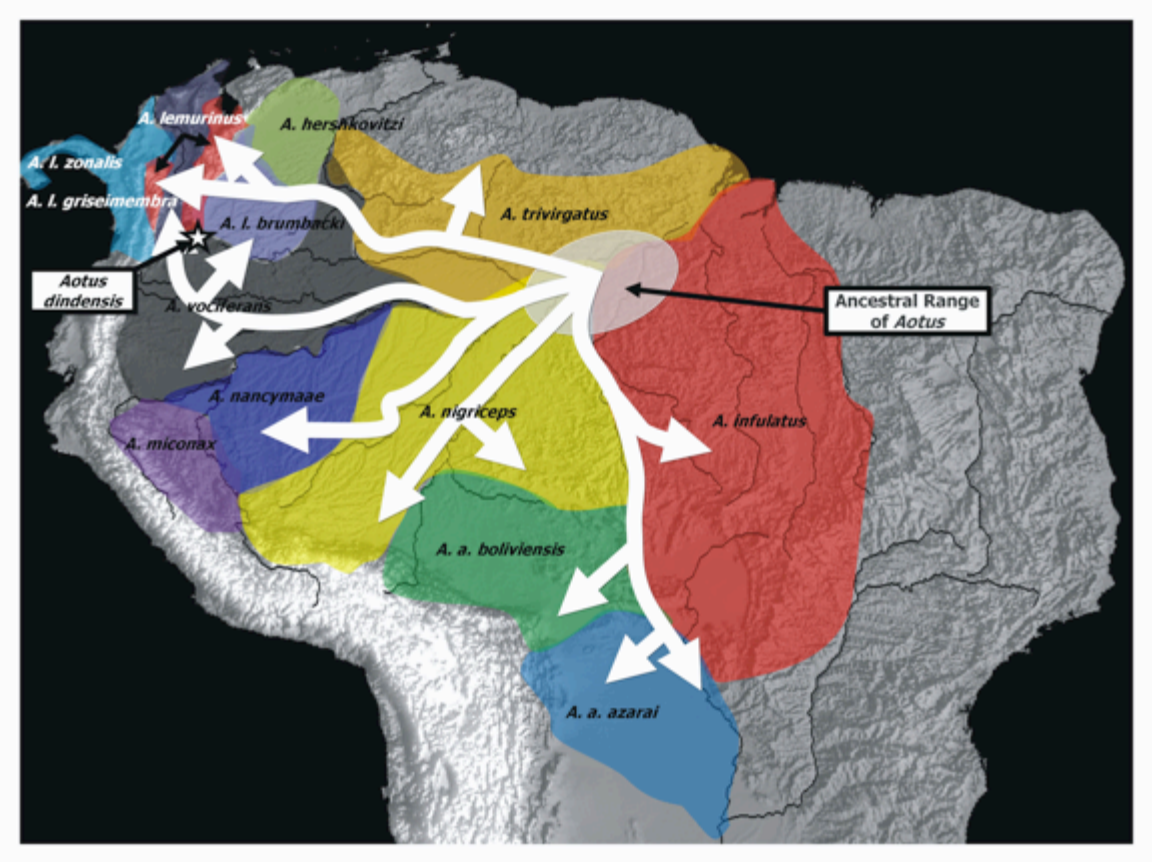


Figure 2.8. Hypothetical scenario for the radiation of the genus *Aotus*. The directionality of the population dispersal is posited based on current geographic distributions of owl monkey species and their genetic relatedness as inferred from mtDNA COII sequences and cytogenetic similarities (**Table 2.1**).

and pelage coat color observed in species from those geographic regions (Ford, 1994; Defler and Bueno, 2007; Fernandez-Duque, 2011).

The southern "red-neck" clade is comprised of species whose COII sequences cluster on the basis of taxon identity (**Figure 2.5a**, **Table 2.1**). Its branches are short and split in a derivative pattern of bifurcations that mimics the same north-south range distribution of those putative taxa. This clinal pattern points to periods of gradual expansion characterized by short episodes of COII sequence divergence among the southern groups. Our coalescence dates also suggest that *A. nigriceps* was the first southern species to diverge from other owl monkey populations beginning some 3.46

mya.

In support of this interpretation, the majority of southern species are karyotypically identical. All of them possess 49(male)/50(female) karyotypes, with the exception of *A. nigriceps* (51m/52f). These findings suggest that the southern expansion of *Aotus* was gradual (e.g., only one chromosomal fission event and the maintenance of the Y-autosomal fusion event in southern males), with species diversifying steadily in different points in time, not through multiple splits or population bottlenecks (Pieczarka et al., 1993, 1998; Torres et al., 1998). This scenario would fit with the paleogeographic history of the South American continent and the Amazon River, including the formation of the gigantic inland Lago Amazonas by the Andean uplift ~9 mya and its drainage ~5.0-2.5 mya, as well as the more recent establishment of southern rivers and the draining of the South American Chaco (Rosenberger et al., 2009).

The northern "grey-neck" clade tells a different evolutionary story. Overall, northern owl monkey taxa display phylogenetic relationships characterized by the separation of several clusters, each with shallow bifurcations. Species found north of the Amazon possess a wide range of karyotypes ranging from 46 in *A. vociferans* to 56 in *A. lemurinus* (Defler and Bueno, 2007). Thus, these two patterns of variation may reflect speciation through neutral genetic drift and isolation through paleogeological river formation, allopatric distance, or the possible consequences of karyotype incompatibility.

Alternatively, some external pressure, such as the gradient of severity of malarial parasitism from the tropics to the subtropics could have been involved in the selection of different pathogen regimes. Such a scenario might explain the high levels of allelic diversity at HLA and KIR loci present in different human populations indigenous to the

same geographic regions (Belich et al., 1992; Parham et al., 1997, Gendzekhadze et al., 2006). If similar pressures were exerted upon owl monkey populations living in the tropics, then it is possible that northern *Aotus* taxa could have experienced more rapid rates of evolution at immunological loci, as well as structurally across their genomes (Van Valen, 1973).

Our data further suggest that *A. nancymaae* was the first to diverge from the northern clade around 6.38 mya, once the north-south split had occurred. Its mtDNA lineage is distinguished by ten characteristic COII nucleotide changes relative to the nearest *Aotus* branch, and it is the only owl monkey taxon to possess a 54(m)/54(f) karyotype. The extent of the molecular distinction of this species suggests the existence of an ancient and temporally pervasive obstacle to gene flow, such as the westward migration of ancestral *A. nancymaae* individuals to western Brazil and eastern Perú, followed by their complete isolation from other owl monkey groups. However, the data and analyses presented in this study do not allow us to distinguish between the effects of population processes and natural selection. Thus, the scenarios described here and above are just a few of the possible interpretations of past events that could have shaped owl monkey evolution.

We also observe that *A. vociferans*, *A. l. griseimembra,* and *A. l. brumbacki* split from *A. lemurinus* and *A. trivirgatus*. This is a curious result, as subspecies *A. l. griseimembra* and *A. l. brumbacki* cluster with the *A. vociferans* clade rather than with the other species of *A. lemurinus*, to which *A. trivirgatus* is closely associated. This branching order may reflect problems with the provenance of these samples or even the accession of numts into databases like GenBank in the place of true mitochondrial

sequences, as was suggested by Menezes et al. (2010). Another explanation for these discrepancies is that, the longstanding disagreement on species definitions for the genus *Aotus,* coupled with the proximity of their ranges to one another, have resulted in the misapplication of species names to samples from the field, or even the incorrect identification of captive animals. Alternatively, a biological influence, such as the hybridism among northern taxa, recently confirmed at the chromosomal level (Monsalve and Defler, 2011), may be responsible for the apparent polyphyly of *A. lemurnius*.

The greater antiquity of northern *Aotus* taxa would have allowed for the many karyotypic changes seen in different owl monkey species. An older time frame for the diversification of the genus (8.95 mya) would also fit the 11.8-13.5 mya estimates for the *Aotus dindensis* fossils. In this regard, it has been argued that their features closely resemble those of all modern owl monkey forms, not just of those species that are geographically nearby (Setoguchi and Rosenberger, 1987; Rosenberger et al., 2009; Takai et al., 2009).

mtDNA diversity and phylogeography of A. a. azarai

The majority of *A. a. azarai* individuals (115/118) belong to one of three major CR haplogroups or clades. Hg-A is the most diverse of the three and, therefore, potentially the oldest. This interpretation is consistent with ρ coalescence estimates for each of the three haplogroups, and is reinforced by the phylogenetic affinity of hg-A to the most distantly related haplotypes X, Y, and Z. The patterning of the *A. a. azarai* mitochondrial networks, along with the general agreement among different summary statistics (D, F_S, τ, and π), reinforces the possibility that this population has undergone

several distinct expansion events in its history, not just a single recent expansion. Even so, it is important to note that the mtDNA genome represents only a single realization of the evolutionary process, and that only the female population history is uncovered by this kind of analysis.

A relatively high degree of CR haplotype sharing among social groups (Fernandez-Duque et al., 2011) limits our ability to reconstruct the colonization processes of the gallery forests along the Pilagá River. However, given that hg-A is the oldest of the three major clades and that it is ubiquitous in most social groups, it is likely to be the ancestral lineage for this population. The relative ages of the derived hg-B and hg-C clades are also consistent with the climatic and geographic processes that drained the southernmost Chacoan forests and flatlands of water some 5,000-7,000 years ago (Iriondo, 1984, 1993). Thus, our haplogroup age estimations suggest the expansion of the population into the Pilagá watershed during, or soon after that period. This scenario would also support the hypothesis that both putative *A. azarai* subspecies, *A. a. boliviensis* and *A. a. azarai*, had common origins further north, and only recently moved southward into the newly accessible South American Gran Chaco.

Chapter 3: Population genetics and kinship in Azara's owl monkeys

Summary[5]

In this study, we characterized a panel of 24 microsatellite markers that reproducibly amplify in Azara's owl monkeys (*Aotus azarai*) for use in genetic profiling analyses. A total of 129 individuals from our study site in Formosa, Argentina, were genotyped for 24 markers, 15 of which were found to be polymorphic. The levels of allelic variation at these loci provided paternity exclusion probabilities of 0.852 when neither parent was known, and 0.984 when one parent was known. In addition, our analysis revealed that, although genotypes can be rapidly scored using fluorescent-base fragment analysis, the presence of complex or multiple STR motifs at a microsatellite locus could present the appearance of equally sized fragment patterns from alleles that have different nucleotide sequences, and perhaps different evolutionary origins. Even so, this collection of microsatellite loci, which is suitable for parentage analyses, will allow us to test various hypotheses about the relationship between social behavior and kinship in wild owl monkey populations. Furthermore, given the limited number of platyrrhine-specific microsatellite loci available in the literature, this STR panel represents a valuable tool for population studies of other cebines and callitrichines.

[5] **The results presented in this chapter have been published in the peer-reviewed report:**
Babb PL, McIntosh AM, Fernandez-Duque E, Di Fiore A, Schurr TG. (2011). An optimized genotyping strategy for assessing genetic identity and kinship in Azara's owl monkeys (*Aotus azarai*). *Folia Primatol 82*, 107-117.

Four additional STRs (D2S95, D5S117, D10S2327, and D16S417) have been added to the panel since the time of the abovementioned publication.

3.1 Introduction

Azara's owl monkeys exhibit a suite of mating and parenting behaviors that is quite rare among primates. They live in small, socially monogamous groups containing a single pair of reproducing adults (Fernandez-Duque, 2011), where the adult male is heavily involved in offspring care, carrying the infant most of the time, and playing, grooming, and even sharing food with it (Rotundo et al., 2005; Wolovich et al., 2006, 2008; Fernandez-Duque et al., 2009). It has long been assumed that *Aotus* males provide care for offspring that they have sired, but this assumption has yet to be formally tested. Thus, we screened a large number of biparentally inherited autosomal microsatellite markers isolated from a range of primate taxa to identify a set of highly variable loci. This set of loci is suitable for delineating genetic identity and assessing parentage in the wild population of *A. azarai* inhabiting our study site in Formosa, Argentina.

Microsatellites, also known as short tandem repeats (STRs), are regions of the genome consisting of sequential repeats (generally 10 or more) of a 2-6 base pair motif that are very likely to differ between any two sampled chromosomes (Pena and Chakraborty, 1994; Morin et al., 1997; Di Fiore, 2003; Kayser et al., 2004). Although numerous microsatellite loci have been characterized in hominoid and cercopithecoid primates, many fail to amplify or display a diminished level of allelic diversity in platyrrhines, which subsequently reduces their efficiency in determining individual identity, parentage, or kinship (Morin et al., 1997, 1998; Witte and Rogers 1999; Lau et al., 2004; Di Fiore and Fleischer, 2004, 2005; Oklander et al., 2007; Raveendran et al., 2008). In addition, chromosomal fissions, fusions, and rearrangements are abundant among platyrrhines, generating variation in their karyotypes even among species of the

same genus (Ma, 1981; Ma et al., 1985; Defler and Bueno, 2003, 2007; Stanyon et al., 2003; Katoh et al., 2009). These facts make it difficult to identify orthologous repetitive regions in disparate *Aotus* taxa and other platyrrhine primates through computational alignment.

As a result, researchers have had to construct platyrrhine-specific genomic libraries of putative STR-bearing vectors, and then sequence and screen these loci in a population sample to identify those sufficiently variable for population genetics studies (Hamilton et al., 1999; Grativol et al., 2001; Böhle and Zischler, 2002; Di Fiore and Fleischer, 2004; Perez-Sweeney et al., 2005; Oklander et al., 2007). Despite the laboriousness of this approach, the isolation of suitable STR markers for a particular population is still subject to chance, and the genomic positions of markers identified in this manner often remain unknown.

Fortunately, an increasing number of genetic studies involving platyrrhine populations have examined these novel microsatellites across platyrrhine taxa (Ellsworth and Hoelzer, 1998; Nievergelt et al., 1998; Witte and Rogers, 1999; Escobar-Paramo, 2000; Böhle and Zischler, 2002; Di Fiore and Fleischer, 2004, 2005; Grativol et al., 2004; Lau et al., 2004; Amaral et al., 2005; Huck et al., 2005; Perez-Sweeney et al., 2005; Oklander et al., 2007; Muniz and Vigilant, 2008). Several of these studies include assessments of individuals from the genus *Aotus*. Building on this work, we developed a panel of STR loci that exhibits sufficient allelic variation to make statistical estimates of kinship between any two individuals in one species of owl monkey (*Aotus azarai*) found at the extreme southern end of the genus' distribution.

3.2 Methods

Samples

Most samples used in this study were collected from 125 wild individuals in >25 social groups from a wild population of Azara's owl monkeys that inhabits the gallery forests along the Pilagá and Guaycolec Rivers in the province of Formosa, Argentina (Fernandez-Duque, 2009). Upon capture, each animal was given a physical exam during which hair, blood, or tissue samples (ear punches, skin biopsies) were collected for use in genetic analyses (Fernandez-Duque and Rotundo, 2003).

High quality DNA was extracted from either tissue or blood using QIAamp DNA Mini and DNA Blood Mini purification kits according to the manufacturer's protocols.

PCR Conditions and Microsatellite Genotyping

Following an extensive literature search, we identified a set of microsatellite loci that were isolated from either platyrrhine primates or humans, and were known to be variable in at least one New World primate species. From this set, we then selected 25 loci that had been noted in previous reports as being polymorphic among *Aotus* individuals. All loci were optimized for PCR amplification through the modulation of annealing temperatures and $MgCl_2$ concentrations on a TouchGene gradient thermocycler (Techne) using "touch-down" PCR (TD-PCR) cycling parameters (Don et al., 1991; Korbie and Mattick, 2008).

Upon optimization, all samples were genotyped for the STR loci using GeneAmp 9700 thermocyclers (ABI). All reactions used AmpliTaq Gold polymerase (ABI) with fluorescent forward primers (5' labeled with 6-FAM dye) and non-fluorescent reverse

primers (Sigma Genosys). DNA samples were diluted to 5 ng/μL working concentrations, from which four μL (20 ng) of DNA were used per PCR reaction (one individual at one microsatellite locus). Although high quality DNA samples were used in every experiment, each genotyping reaction was repeated at least four times to confirm presence of homozygotes (Morin et al., 2001; Soulsbury et al., 2009). The specific PCR reagent volumes and cycling conditions for the different loci are listed in **Table 3.1**.

Table 3.1. Primers and PCR conditions.

Locus	Identified in Species	Repeat motif	Primer	Primer sequence (5'-3')	Size range (bp)	PCR	Variable in *Aotus*
1115	*Lagothrix lagotricha*[a]	$(GT)_3(GA)(GT)_5(CT)(GT)_{11}$	1115-F	6*FAM-GCTCATATTCATACATCCCTTGG	179 - 193	45cyc TD (62-57)	Yes
			1115-R	TTTGCTTGCTCATTCATTGC			
1118	*Lagothrix lagotricha*[a]	$(GA)_2(GT)_{10}$	1118-F	6*FAM-TTTCTCCCTCTCAGATTACCAG	172	40cyc TD (57-52)	No
			1118-R	CCTTGAGGTTTTTGGGTTCC			
113	*Lagothrix lagotricha*[a]	$(GT)_{15}$	113-F	6*FAM-GCAAAACTCCCCTGTGACTG	143	40cyc TD (57-52)	No
			113-R	CCCACTCTCCTCCACAAAGG			
Ap40	*Alouatta palliata*[b]	$(TG)_4CA(TG)_6$	Ap40-F	6*FAM-CCACGGTGGCAGAGGAGATTT	171	40cyc TD (59-54)	No
			Ap40-R	AGAGGCACGAAGACAAGGACA			
Ap68	*Alouatta palliata*[b]	$(TG)_{17}$	Ap68-F	6*FAM-TGTTGGTATAATCTTTCCTA	164	40cyc TD (57-52)	No
			Ap68-R	ACATACACCTTTGAGTTTCT			
Ap74	*Alouatta palliata*[b]	$(TG)_{19}$	Ap74-F	6*FAM-TGCACCTCATCTCTTTCTCTG	146	40cyc TD (57-52)	No
			Ap74-R	CATCTTTGTTTTCCTCATAGC			
CJ13	*Callithrix jacchus*[c]	$(CA)_{13}(TC)_{14}$	CJ13-F	6*FAM-CAAGGAAACATAAGTGTGGCTC	238	40cyc TD (60-50)	No
			CJ13-R	CAGACATTTTAGCCCCTTCC			
CJ14	*Callithrix jacchus*[c]	$(TC)_{21}(CA)_5$	CJ14-F	6*FAM-CTTGTGACAGTGGGGGAGTT	124 - 158	40cyc TD (66-56)	Yes
			CJ14-R	CAAGTGTGAACATCCATGCC			

Locus	Identified in Species	Repeat motif	Primer	Primer sequence (5'-3')	Size range (bp)	PCR	Variable in *Aotus*
D10S2327	*Isolated from Homo sapiens and screened in Aotus spp.*[d]	$(TGGA)_{10}(CGGA)(TGGA)(CGGA)_2$	D10S2327-F	6*FAM-CCCAGAGCAAGTACTCACCT	182	45cyc TD (60-55)	No
			D10S2327-R	ATAGTTTTGTGCTTATTGACATGA			
D13S160	*Isolated from Homo sapiens and screened in Aotus spp.*[d]	$(AT)_{10}$	D13S160-F	6*FAM-CGGGTGATCTAAGGCTTCTA	220 – 242	40cyc TD (60-50)	Yes
			D13S160-R	GGCAGAGATATGAGGCAAAA			
D15S108	*Isolated from Homo sapiens and screened in Aotus spp.*[d]	$(CT)_5(AT)_5$	D15S108-F	6*FAM-AGGAGAGCTAGAGCTTCTAT	170 - 186	40cyc TD (56-51)	Yes
			D15S108-R	GTTTCAACATGAGTTTCAGA			
D16S417	*Isolated from Homo sapiens and screened in Aotus spp.*[d]	$(CA)_{15}(TA)(CA)_2$	D16S417-F	6*FAM-CTGTCCAACATGCAGCC	128-134	45cyc TD (60-55)	Yes
			D16S417-R	TGAAGTCAATCCCACTTGAA			
D2S95	*Isolated from Homo sapiens and screened in Aotus spp.*[d]	$(TG)_{17}$	D2S95-F	6*FAM-GACAGAGCAACACCCCAACT	154	45cyc TD (60-55)	No
			D2S95-R	TCATCACTCACCCAGACCAA			
D4S411	*Isolated from Homo sapiens and screened in Aotus spp.*[d]	$(TA)_{20}$	D4S411-F	6*FAM-AGGCTGTCTTGGCAGAAAT	134-168	40cyc TD (60-50)	Yes
			D4S411-R	GATGTAATCCTGTGCTATGGC			
D5S117	*Isolated from Homo sapiens and screened in Aotus spp.*[d]	$(AC)_{17}$	D5S117-F	6*FAM-TGTCTCCTGCTGAGAATAG	128-146	45cyc TD (60-55)	Yes
			D5S117-R	TAATATCCAAACCACAAAGGT			
D8S275	*Isolated from Homo sapiens and screened in Aotus spp.*[d]	$(TC)_{20}$	D8S275-F	6*FAM-AAATCGCTAGAAAATGTCCA	137 - 151	40cyc TD (65-60)	Yes
			D8S275-R	TCACACCTGGGAATTAGAAG			
Leon30c73	*Leontopithecus chrysopygus*[e]	$(TC)_{25}(AA)(TC)(TG)_{16}$	Leon30c73-F	6*FAM-GGACCTGATTGAAGCAGTC	240-270	40cyc TD (60-50)	Yes
			Leon30c73-R	TTCCCTGAGAATCTAATGGAG			
PEPC3	*Cebus apella*[f]	$(GT)_{13}$	PEPC3-F	6*FAM-CATGGACTGCAATTCAAGCC	211 - 289	40cyc TD (63-58)	Yes
			PEPC3-R	ACTTCCAGCCTCCAAAACTATG			
PEPC40	*Cebus apella*[f]	$(CA)_{18}(CT)_{14}(CA)_9$	PEPC40-F	6*FAM-GACAGAGCAAGACTCCATCTC	160 - 164	40cyc TD (55-50)	Yes
			PEPC40-R	GATCAGTAAACACATGTGCAT			
PEPC59	*Cebus apella*[f]	$(GT)_{18}$	PEPC59-F	6*FAM-CAGTGGCAACTCTGTAAGGA	256	40cyc TD (59-54)	No

Locus	Identified in Species	Repeat motif	Primer	Primer sequence (5'-3')	Size range (bp)	PCR	Variable in *Aotus*
PEPC8	*Cebus apella*[f]	$(CA)_{16}$	PEPC59-R	GTGGAGTCAACATGCAGAGG	257 - 259	40cyc TD (63-58)	Yes
			PEPC8-F	6*FAM-TTCAGGATGCATCAAATGATT			
			PEPC8-R	TAGCAGTCTATTTAGGTGTTAAT			
SB24	*Saguinus bicolor*[g,h]	$(CA)_{23}$	SB24-F	6*FAM-ATCTGCCTATCACTTCTTTC	118 - 138	40cyc TD (59-54)	Yes
			SB24-R	CATTTGCTCTGCTCATTCA			
SB38	*Saguinus bicolor*[g,h]	$(CA)_{19}$	SB38-F	6*FAM-GCCTCAATGGGTTTTAACC	89 - 145	40cyc TD (55-50)	Yes
			SB38-R	AGAACGAGTCTGTATCTTGA			
SW34D	*Saimiri boliviensis*[i]	$(CA)_{14}$	SW34D-F	6*FAM-CATCAAAGGATATTATTATC	117 - 143	40cyc TD (61-56)	Yes
			SW34D-R	TACATTTCTGGATACTAGGC			
SW65B	*Saimiri boliviensis*[i]	$(AT)_5GT(AT)_5(AC)_{18}$	SW65B-F	6*FAM-TGAAGTAATAAAATACATAG	n/a	n/a	n/a
			SW65B-R	ACATTAGGGTCGATGAGTCC			

Data sources: [a] Di Fiore and Fleischer, 2004, [b] Ellsworth and Hoelzer, 1998, [c] Nievergelt et al., 1998, [d] Lau et al., 2004, [e] Perez-Sweeney et al., 2005, [f] Escobar-Paramo, 2000, [g] Böhle and Zischler, 2002, [h] Huck et al., 2005, [i] Witte and Rogers, 1999

PCR amplification success was evaluated via standard gel electrophoresis using high resolution 3% NuSieve agarose (Lonza) and digitally recorded using the Molecular Imaging v4.5.1 software on a Kodak GL200 Imaging Station (Carestream Health). Aliquots of successful PCR amplifications were subsequently combined with 0.3 µL LIZ500 allelic marker standard (ABI) and 8.0 µL Hi-Di formamide (ABI) to make 10 µL final reaction volumes for capillary-based fluorescent fragment analysis. Reactions were denatured at 95°C for three minutes, and then immediately placed on ice for three minutes. All reactions were run on a 3130xl Gene Analyzer (ABI) for fragment analysis, and alleles were read and scored using GeneMapper ID v3.2 software (ABI).

DNA Sequencing of Microsatellites

Individual alleles of the different STR loci were amplified using unlabelled versions of the same primers from homozygous individuals, and then sequenced in order to identify their repeat motif and repeat length in *A. azarai*. Amplicons were first purified by SAP/Exo I digestion (New England BioLabs) and cycle-sequenced using Big Dye™ Terminator v3.1 (ABI) sequencing chemistry. Excess dye terminators were removed with the BigDye XTerminator™ purification kit (ABI), and the cycle-sequencing products were separated via capillary electrophoresis on a 3130xl Gene Analyzer (ABI). Read quality of chromatograms was assessed using Sequencing Analysis v5.4 software (ABI), and bidirectional sequences were aligned and assembled using Sequencher v4.9 (Gene Codes) and Geneious Pro v5.5.5 (Drummond et al., 2010). All sequences generated in this study have been uploaded to NCBI's GenBank nucleotide database (accession numbers JN609277-JN609289).

Statistical and Genomic Analyses

Molecular summary statistics were computed using the population genetic software package Arlequin v3.11 (Excoffier et al., 2005). Individual identity analyses, parentage exclusion probabilities and locus-specific heterozygosity indices were calculated using the software CERVUS v3.0.3 (Marshall et al., 1998; Slate et al., 2000).

The physical locations of the STR loci were estimated using the sequenced DNA templates generated from our *A. azarai* samples with the BLAT algorithm on the UCSC Genome Browser to interrogate marmoset (calJac3), rhesus macaque (rheMac2), and human (hg18) chromosomes (Kent et al., 2002; Karolchik et al., 2008). This analysis

was undertaken to identify any two loci that might be genetically linked based on their chromosomal proximity to one another. The chromosomal locations were selected based on nucleotide similarity (>95%) of the entire sequenced amplicons to these different primate genomes.

3.3 Results

Of the 25 STR loci that we tested, 24 were optimized for PCR amplification and fragment analysis of *A. azarai* samples, resulting in 48 reliable allele calls for each individual. Of these 24 STR loci, 9 were monomorphic and 15 were polymorphic in the 129 individuals that were genotyped. The panel of 15 polymorphic markers averaged 6.07 alleles per locus. The mean percentage of the 129 individuals genotyped per locus was 97.31%. The average heterozygosity for the loci in this panel was 0.4301, and the panel exhibited a mean polymorphic information content value (PIC) of 0.3821. Locus-specific values are listed in **Table 3.2**.

Table 3.2. Allelic variation and heterozygosity indices for 15 loci from the *Aotus azarai* panel.

Locus	No. of Alleles	# Typed	H_O	H_E	HWE (*P*)	PIC	Null Allele Freq (F)	Sequence Accession #
1115	2	123	0.195	0.177	NS	0.161	-0.045	JN609277
CJ14	7	127	0.402	0.441	NS	0.407	0.029	JN609278
D13S160	9	128	0.836	0.566	<0.0000001	0.469	-0.208	JN609279
D15S108	4	129	0.349	0.361	NS	0.326	0.019	JN609280
D16S417	6	125	0.456	0.488	NS	0.447	0.023	-
D4S411	10	129	0.829	0.733	0.0004	0.685	-0.076	JN609281
D5S117	8	128	0.609	0.63	NS	0.567	0.016	-
D8S275	7	129	0.395	0.379	NS	0.358	-0.041	JN609282
Leon30c73	5	124	0.177	0.198	NS	0.183	0.063	JN609283
PEPC3	4	128	0.414	0.337	NS	0.285	-0.101	JN609284
PEPC40	2	129	0.326	0.397	NS	0.317	0.097	JN609285
PEPC8	2	129	0.395	0.463	NS	0.355	0.077	JN609286
SB24	8	127	0.205	0.218	NS	0.209	0.014	JN609287
SB38	12	100	0.33	0.652	<0.0000001	0.628	0.347	JN609288
SW34D	5	128	0.555	0.412	0.0002	0.336	-0.150	JN609289

Note: The STR loci 1118, 113, Ap40, Ap68, Ap74, CJ13, D2S95, D10S2327 and PEPC59 were monomorphic (1 allele present) in *Aotus azarai.* In addition, we were unable to reproduce results from PCR and fragment analysis for locus SW65B in *Aotus azarai*.
H_O: Observed heterozygosity
H_E: Expected heterozygosity
HWE (*P*): Significance (*P*) in the deviation from Hardy-Weinberg equilibrium
PIC: Polymorphic information content

The observed heterozygosity values (H_O) for each locus were compared to their respective expected heterozygosity estimates (H_E), and the genotype frequencies for most of the loci (11 of 15) did not significantly deviate from those expected under Hardy-Weinberg equilibrium. Of the four loci that deviated from Hardy-Weinberg equilibrium, D4S411, D13S108, and SW34D had higher heterozygosity values than expected, whereas SB38 exhibited lower heterozygosity values than expected. In addition, by using CERVUS to estimate null allele frequencies (F) at each locus, we calculated that the predicted occurrence of null alleles ranged from -0.21 to +0.35 across the different markers.

When combined, the 15 polymorphic loci provided a parentage exclusionary percentage of 85.2% when both parents were unknown, and 98.4% when one parent was unknown. These percentages increased to 99.9% when both parents were known, 100%

for the detection of individual identity, and 99.9% for the estimation of sibling identities. However, at the three loci with complex repeat motifs (D4S411, D8S275, Leon30c73, PEPC40, and SW34D), we observed that alleles that had been genotyped as being identical (i.e., equal PCR fragment lengths) possessed different DNA sequences, suggesting the possibility of "size homoplasy" at these markers (Estoup et al., 1995; Viard et al., 1998; Roeder et al., 2009). The complexity of repeat motifs present at these five loci is illustrated in **Figure 3.1**.

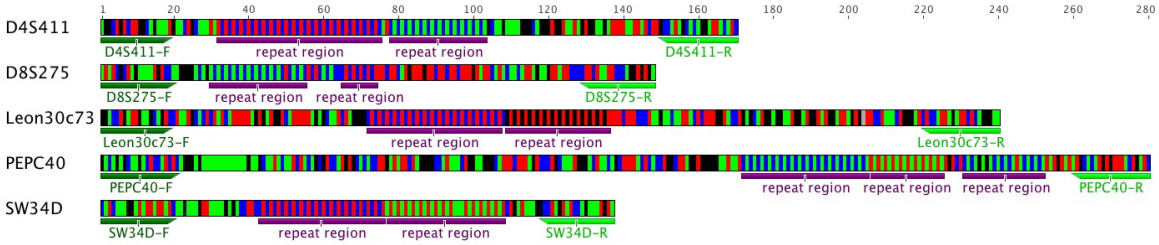


Figure 3.1. A schematic diagram of five "complex" STR loci in *A. a. azarai* showing the composition of multiple repetitive elements found within each amplicon (2-3 motifs per locus). Arrows indicate the position and direction of the oligonucleotide primers. These sequences of these primers, along with detailed PCR recipes and cycling parameters, are listed in **Table 3.1**.

We did not detect any significant linkage disequilibrium among any loci in the panel. To confirm this observation, we mapped the physical location of each locus by comparing its nucleotide sequence to the genomes of other primates. The BLAT results matched the loci in our panel to a minimum of 11 different marmoset chromosomes, 12 different rhesus macaque chromosomes, and 12 different human chromosomes (**Table 3.3**).

Table 3.3 Chromosomal localization of STR loci comprising the *Aotus azarai* panel

Locus	Identified in Species	Marmoset (calJac3)	Rhesus macaque (rheMac2)	Human (hg18)
1115	*Lagothrix lagotricha*	Chromosome 7	Chromosome 9	Chromosome 10

Locus	Identified in Species	Marmoset (calJac3)	Rhesus macaque (rheMac2)	Human (hg18)
1118	*Lagothrix lagotricha*	Chromosome 12	Chromosome 9	Chromosome 10
113	*Lagothrix lagotricha*	Chromosome 12	Chromosome 20	Chromosome 16
Ap40	*Alouatta palliata*	Chromosome 1	Chromosome 17	Chromosome 13
Ap68	*Alouatta palliata*	Chromosome 12	Chromosome 20	Chromosome 16
Ap74	*Alouatta palliata*	Chromosome 3	Chromosome 5	Chromosome 4
CJ13	*Callithrix jacchus*	Chromosome 11	Chromosome 14	Chromosome 11
CJ14	*Callithrix jacchus*	-	-	-
D13S160	*Homo sapiens*	Chromosome 1	Chromosome 17	Chromosome 13
D15S108	*Homo sapiens*	Chromosome 10	Chromosome 7	Chromosome 15
D16S417	*Homo sapiens*	Chromosome 12	Chromosome 20	Chromosome 16
D4S411	*Homo sapiens*	Chromosome 3	Chromosome 5	Chromosome 4
D5S117	*Homo sapiens*	Chromosome 2	Chromosome 6	Chromosome 5
D8S275	*Homo sapiens*	Chromosome 16	Chromosome 8	Chromosome 8
Leon30c73	*Leontopithecus chrysopygus*	Chromosome 6	Chromosome 12	Chromosome 2
PEPC3	*Cebus apella*	Chromosome 22	Chromosome 19	Chromosome 19
PEPC40	*Cebus apella*	Chromosome 14	Chromosome 13	Chromosome 2
PEPC59	*Cebus apella*	Chromosome 6	Chromosome 7	Chromosome 15
PEPC8	*Cebus apella*	Chromosome 4	Chromosome 4	Chromosome 6
SB24	*Saguinus bicolor*	Chromosome 3	Chromosome 5	Chromosome 4
SB38	*Saguinus bicolor*	Chromosome 12	Chromosome 20	Chromosome 16
SW34D	*Saimiri boliviensis*	-	-	-

Data Sources: BLAT alignment program (Kent et al., 2002) and the UCSC Genome Browser (Karolchik et al., 2008).

Note: The highly repetitive sequences for CJ14 and SW34D failed to align to a singular location for any of the genome assemblies examined. Instead, when 95% sequence identity thresholds in BLAT were applied, these two loci mapped to many different genomic locations at low confidence levels.

3.4 Discussion

Our STR optimized panel provides a novel assay for investigating genetic relatedness and kinship in owl monkeys. Our observations from the field have shown that, although male owl monkeys invest significant resources in parenting (Rotundo et al., 2005; Wolovich et al. 2008), it is possible that these resources may be directed at offspring whom they did not sire (Fernandez-Duque et al., 2008; Fernandez-Duque et al.,

2009). Such a finding would represent a clear deviation from the predictions of kin selection and parental investment theories (Hamilton, 1964a, b; Trivers, 1972).

However, testing such scenarios requires the identification of the biological relationships between individuals as delineated by their genetic sequences. Until recently, the lack of platyrrhine STR markers made it difficult to assess genetic relatedness in closely related New World primates, where significant homozygosity often impedes the delineation of individual identity (e.g., Lau et al., 2004). We have shown here that this panel of 15 variable STRs can estimate paternity and kinship with a high degree of certainty among owl monkey individuals, thereby permitting further exploration of the social structure of this monogamous taxon.

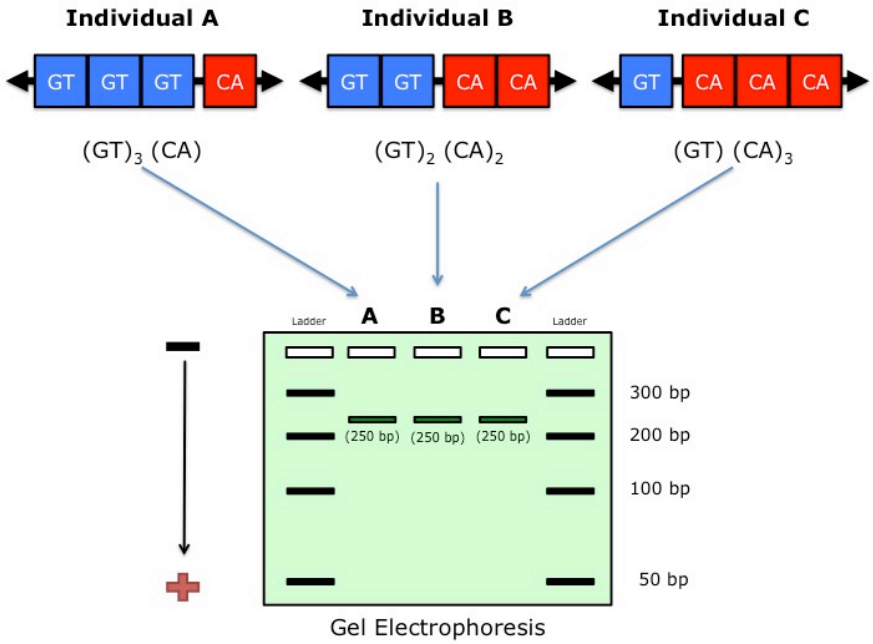


Figure 3.2. A schematic diagram illustrating the potential effects of "size homoplasy" at a hypothetical "complex" STR locus. Within this locus, two repeat motifs (GT and CA) exist, and each of them is variable within the population. In this haploid example, three individuals (A, B, and C) each possess differing numbers of repeat cassettes at each of the two repeat motifs. While the repeat content should distinguish each individual from one another on the basis of sequence composition, they all present equally sized allelic fragments (250 bp) when viewed on a gel (as well as through capillary-based fluorescent fragment analysis). These traditional methods for genotyping based on fragment length comparisons are not always appropriate for the analysis of complex STRs, as all three individuals in this example would all be incorrectly scored as having the same ancestral allelic states.

However, complex STR repeat motifs should be carefully considered in any parentage assessment, as "size homoplasy" could skew the delineation of parental alleles from the interpretation of fragment size alone. For example, the amplification of a locus that possesses two or more distinct repeat motifs, such as Leon30c73 (CT_{18}+GT_{13}), could potentially generate allelic fragments of the same length in individuals that have different combinations of the two repeat motifs (**Figure 3.2**). Occurrences of such "size homoplasies" have been noted in other STR panels that take advantage of cross-specific amplification of microsatellite loci (Roeder et al., 2009; Soulsbury et al., 2009). For this reason, we recommend the direct sequencing of these regions when they are to be used for pedigree or population studies. Nevertheless, our microsatellite panel possesses sufficient exclusionary power for the discernment of individual identity among *Aotus* individuals. It may further prove to be valuable for population and sociobehavioral studies of other cebine and callitrichine taxa.

Chapter 4: *AVPR1A* Variation

Summary[6]

The arginine vasopressin V1a receptor gene (*AVPR1A*) has been implicated in increased partner preference and pair bonding behavior in mammalian lineages. This observation is of considerable importance for studies of social monogamy, which only appears in a small subset of primate taxa, including the Argentinean owl monkey (*Aotus azarai*). Thus, to investigate the possible influence of *AVPR1A* on the evolution of social behavior in owl monkeys, we sequenced this locus in 25 individuals from our study population from the Gran Chaco. We also assessed the interspecific variation of *AVPR1A* in 23 other primate and rodent species that represent a set of phylogenetically and behaviorally disparate taxa. The resulting data revealed *A. azarai* to have a unique genic structure for *AVPR1A* that varies in coding sequence and microsatellite repeat content relative to other primate and mammalian species. Specifically, one repetitive region that has been the focus in studies of human *AVPR1A* diversity, "RS3", is completely absent in *A. azarai* and all other platyrrhines examined. This finding suggests that, if *AVPR1A* modulates behavior in owl monkeys and other neotropical primates, it does so independent of this region. These observations have also provided clues about the process by which the range of social behavior in the Order Primates evolved through lineage-specific neurogenetic variation.

[6] **The results presented in this chapter have been published in the peer-reviewed report:**
Babb PL, Fernandez-Duque E, Schurr TG. (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.

4.1 Introduction

Primate species exhibit an extensive range of social behaviors and, as a result, display considerable diversity in their social organization and reproductive strategies, including social monogamy. Owl monkeys (*Aotus* spp.), saki monkeys (*Pithecia* spp.) and titi monkeys (*Callicebus* spp.) exemplify the minority of primate species living in small social groups (Kleiman, 1977; Wright, 1994; Van Schaik and Kappeler, 2003). These particular platyrrhines, or New World monkeys, along with other taxa such as the indris, fat-tailed dwarf lemurs, callitrichids and gibbons, are traditionally defined as possessing prolonged and essentially exclusive mating relationship between one male and one female, which comprise the basic social unit (Kleiman, 1977; Mock and Fujioka, 1990; Palombit, 1994). Such monogamous social systems are rare among primates, but are widely distributed across the order, indicating that they have arisen independently multiple times (van Schaik and van Hoof, 1983; van Schaik and Dunbar, 1990). Although "monogamy" necessarily involves many different complex behaviors (Mendoza et al., 2002; Moller, 2003), this scenario raises an important question: are the social behaviors of primates controlled by the same genes in different species? By comparing genetic regions that are related to these behaviors across primate taxa, one can begin to determine whether the evolution of complex social behavior occurs through conserved molecular mechanisms.

Current research on voles (*Microtus* spp.) and other mammals suggests that variation at key neurogenetic loci likely played an important role in the evolution of social behavior at both inter- and intraspecific levels (Lim et al., 2004a). In voles and other rodents, the hormone arginine vasopressin (AVP) and the V1a receptor protein

(V1aR, encoded by the gene *avpr1a* [rodents] / *AVPR1A* [primates]) are implicated in the variable expression of pair bonding in both sexes. This is a necessary behavioral component for the manifestation of monogamous social systems (Young et al., 1999; Hammock and Young, 2002; Turner et al., 2010). However, AVP has directly been linked to a variety of crucial processes in mammals, ranging from the regulation of circadian rhythms, fluid retention, and the vasoconstriction of blood vessels (Thibonnier et al., 1996). The fact that AVP is pleiotropic in its effects complicates efforts to link sequence variation at this locus to behavioral phenotypes.

Variation in the expression of *AVPR1A* and localization of V1aR proteins also influences the manifestation of pair bonding in mammalian species. The differential neuroanatomical distribution of V1aR G-protein-coupled receptor (GPCR) proteins functionally distinguishes monogamous prairie voles (*Microtus ochrogaster*) from their polygynous relatives (Pitkow et al., 2001). In addition, viral vector-mediated transfers of the *AVPR1A* gene into the ventral forebrain of polygynous meadow voles (*Microtus pennsylvanicus*) demonstrated a significant increase in their likelihood to exhibit a partner preference (Lim et al., 2004a, b), while other studies using *in vitro* expression assays have revealed that inter- and intraspecific differences in short tandem repeat (STR) length influence *AVPR1A* expression (Hammock and Young, 2004, 2005). Differences in V1aR density and localization in the brains of different mammalian taxa are also associated with specific lengths of microsatellite STRs found in the 5' upstream promoter region *AVPR1A*, and are positively correlated with the expression of the gene in a tissue-specific manner (Hammock and Young, 2002, 2004, 2005; Hammock et al., 2005; Young and Hammock, 2007; Young et al., 2005). Overall, these studies demonstrate that

AVPR1A/*avpr1a* expression influences the distribution of V1aR in the brain, which ultimately influences social behavior.

These observations suggest that, because regulatory sequences drive the expression of the gene, variation in and around these regions may be responsible for emergence of different social structures in different species. Nevertheless, despite their presence in at least four distinct clusters in the 5' flanking region, STR patterns are not the only component responsible for differences in brain expression of *AVPR1A*. As noted by Fink et al. (2006), STR patterns represent more of a phylogenetic oddity than a consistent pattern related to monogamy versus polygamy. Instead, amino acid changes in the coding region of *AVPR1A* may explain sociobehavioral differences across species (Fink et al., 2007; Phelps et al., 2009; Turner et al., 2010).

In this regard, recent work tracing the sequence evolution of *AVPR1A* in catarrhine and hominoid primates has revealed both intra- and interspecific levels of molecular variation (Donaldson et al., 2008; Rosso et al., 2008). These differences consist of numerous amino acid changes that have accumulated in different taxa, as well as lineage-specific STR motifs and length distributions (**Figure 4.1**). It has been suggested that different aspects of these variable components in or surrounding *AVPR1A* have been involved in the evolution of different primate social systems (Donaldson et al., 2008; Rosso et al. 2008). However, despite the diversity of social behaviors observed in platyrrhine primates, including the apparent enrichment of monogamous social systems, there has only been one study of *AVPR1A* sequence variation in these species (Babb et al., 2010).

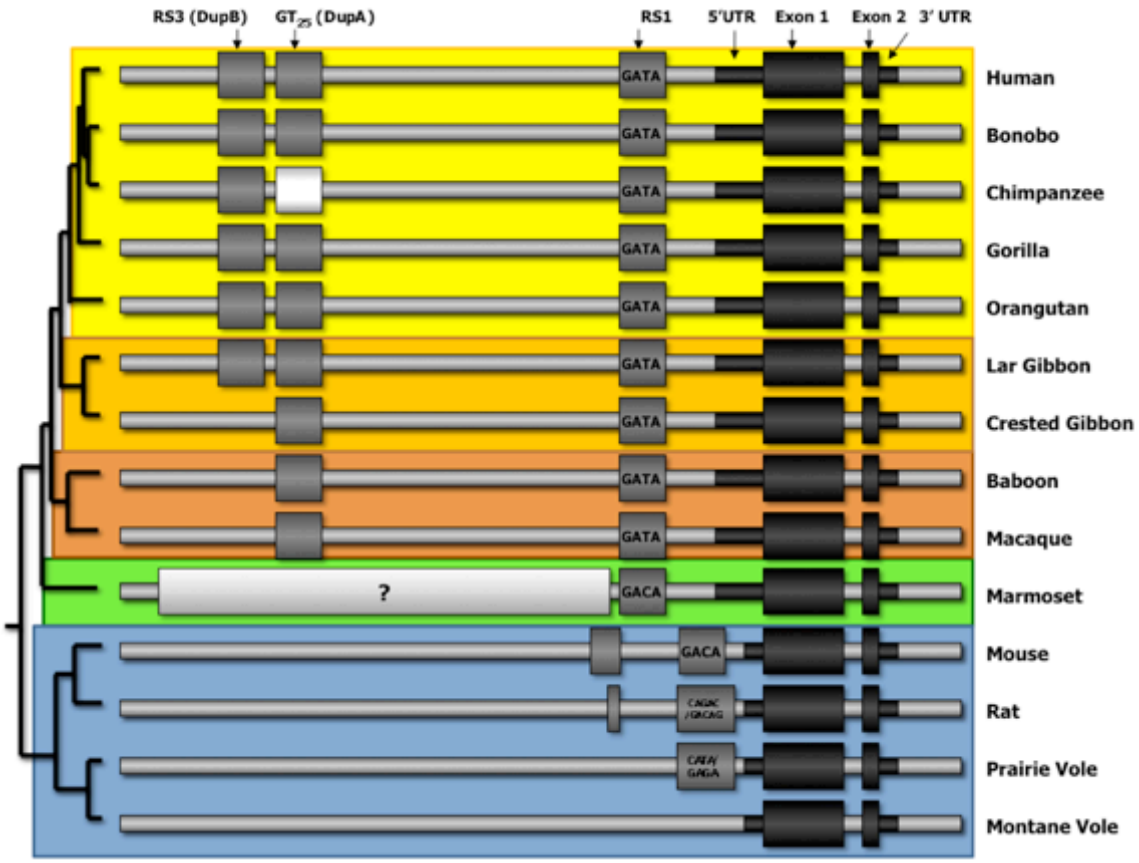


Figure 4.1. A phylogenetic representation of the genic structure of the *AVPR1A* locus in different mammalian species. Monogamous voles are known to possess a singular repeat motif known as Repeat Sequence 1 (RS1) of CATA/GAGA that is located roughly 600 nucleotides upstream of the TSS (Hammock and Young, 2005). In mice and rats, this changes to a CAGAC/GACAG repeat pattern, and rats even have a slight duplication of the region further upstream (Murasawa et al., 1995; Insel, 2003). Although sequence conservation of the *AVPR1A* 5' flanking region on the nucleotide level diminishes with phylogenetic distance, this overall pattern persists in and among primate species. Marmosets exhibit a GACA repeat, and Old World primates possess a GATA repeat at the RS1 location. In Old World monkeys, additional repeat motifs have accumulated, such as the GT_{25} (DupA) repeat motif in macaques and baboons. By contrast, hominoid apes, such as chimps, bonobos, orangutans, and humans, have further developed a duplication of this GT_{25}, denoted in the literature as RS3 (DupB), which is a more complex variable nucleotide tandem repeat (VNTR) pattern. For example, humans have three repetitive sequences in the *AVPR1A* 5' flanking region: a $(GATA)_{14}$ at RS1, a $(GT)_{25}$ dinucleotide repeat, and a complex $(CT)_4$-TT-$(CT)_8$- $(GT)_{24}$ repeat at RS3 (Figure compiled and adapted from Thibbonier et al., 1996; Hammock and Young, 2002, 2003, 2005; Insel, 2003; Rosso et al., 2008; Hammock et al., 2005; Donaldson et al., 2008; Walum et al., 2008). The different taxa compared here are displayed on colored backgrounds corresponding to their phylogenetic placement. *Yellow* large bodied apes, *gold* small bodied apes, *orange* Old World monkeys, *green* New World monkeys, *blue* rodents.

To expand on our earlier work, we began exploring the molecular evolution of *AVPR1A* in New World primates to clarify its possible role in the evolution of their mating behavior and social organization. We sequenced the *AVPR1A* locus in our study

population of owl monkeys to characterize the genetic variation present in the 5' promoter and coding region at the intraspecific level. We then sequenced *AVPR1A* in numerous Old and New World primate taxa exhibiting a range of mating systems to assess interspecific levels of genetic variation. We anticipated that the *AVPR1A* gene of owl monkeys and other platyrrhines would exhibit species-specific coding region substitutions relative to other primate taxa, and that these variants could lead to novel structural characteristics of their V1aR proteins. In addition, intraspecific variation of STRs present in the 5' region of the owl monkey *AVPR1A* were expected to exhibit a normally-distributed range of allele sizes that would be commensurate with those seen in previous genotyping surveys of primate populations (Donaldson et al., 2008; Rosso et al., 2008).

Our results reveal novel patterns of genetic variation among platyrrhine primates, and show that this variation distinguishes them from their Old World relatives. These findings suggest that if monogamy has indeed evolved independently in primates on both sides of the Atlantic, then it has done so through different molecular mechanisms.

4.2 Methods

Study population

Azara's owl monkeys, inhabit the gallery forests and the patchy forest-islands interspersed throughout the savannahs of northeastern Argentina. Their social groups are small (two to six individuals), being comprised of an adult heterosexual pair, one infant, and one or two juveniles, and occasionally a subadult (Rotundo et al., 2000). The adult male in the group is heavily involved in the care of the young, and carries the infant most

of the time (84%) after the first week of life (Rotundo et al., 2005). This pattern of social monogamy and paternal care distinguishes owl monkey males from those of the majority of other primate species.

Samples

We assembled a panel of 25 *A. azarai* individuals for the analysis of intraspecific variation of *AVPR1A* regulatory and coding sequences (**Table 4.1**). The *A. azarai* individuals in this panel represent the most disparate mtDNA (maternal) haplotypes in a wild study population (n=118) living in Formosa Province, Argentina (Babb et al., 2010). Upon capture, each of the 20 wild individuals underwent a thorough physical exam (Fernandez-Duque and Rotundo, 2003) during which we obtained biological samples for use in genetic analyses. For one individual, the source of DNA was the remains of a placenta and fetus found on the ground in the savannah ("AaaPLunk"). In addition, we collected four samples from captive individuals at the Saenz-Peña Municipal Zoo, which is located 250 km away from the study area in the city of Saenz-Peña, Chaco Province. The exact geographic origin of these animals is unknown.

For comparative analyses of genetic variation at the *AVPR1A* locus, DNA samples from three putative *Aotus* species and subspecies (*A. nancymaae*, *A. lemurinus*, and *A. l. grisiemembra*) from the Zoological Society of San Diego. Another six *Aotus* samples (*A. nancymaae*, n=5; *A. vociferans,* n=1) were obtained from individuals at the DuMond Conservancy for Primates and Tropical Forests (Miami, FL). In addition, we analyzed samples from several other platyrrhine taxa, including two individuals each from saki

Table 4.1. Samples investigated at the *AVPR1A* locus.

ID	Species	Common Name	Sex	Locale
AaaPLunk[a]	*Aotus azarai*	Azara's Owl Monkey	Unk	Core area, Formosa, AR
Aaa008	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa014	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa015	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa021	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa032	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa034	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa037	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa053	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa057	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa063	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa067	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa071	*Aotus azarai*	Azara's Owl Monkey	M	Downstream, Formosa, AR
Aaa082	*Aotus azarai*	Azara's Owl Monkey	M	Downstream, Formosa, AR
Aaa087	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa092	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa108	*Aotus azarai*	Azara's Owl Monkey	F	Upstream, Formosa, AR
Aaa109	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa114	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa122	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa123	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
AaaF1	*Aotus azarai*	Azara's Owl Monkey	F	Saenz-Pena Zoo, AR
AaaF1B	*Aotus azarai*	Azara's Owl Monkey	F	Saenz-Pena Zoo, AR
AaaF2	*Aotus azarai*	Azara's Owl Monkey	F	Saenz-Pena Zoo, AR
AaaM2	*Aotus azarai*	Azara's Owl Monkey	M	Saenz-Pena Zoo, AR
Al001	*Aotus lemurinus*	Grey-bellied Owl Monkey	Unk	San Diego Zoo/CRES
Alg002	*Aotus lemurinus grisiemembra*	Grey-handed Owl Monkey	Unk	San Diego Zoo/CRES
Ana001	*Aotus nancymaae*	Nancy Ma's Owl Monkey	Unk	San Diego Zoo/CRES
Ppp001	*Pithecia pithecia*	White-faced Saki Monkey	M	San Diego Zoo/CRES
Ppp002	*Pithecia pithecia*	White-faced Saki Monkey	F	San Diego Zoo/CRES
Sss001	*Saimiri sciureus*	Common Squirrel Monkey	M	San Diego Zoo/CRES
Sss002	*Saimiri sciureus*	Common Squirrel Monkey	F	San Diego Zoo/CRES
Ana002	*Aotus nancymaae*	Nancy Ma's Owl Monkey	F	DuMond Conservancy
Ana003	*Aotus nancymaae*	Nancy Ma's Owl Monkey	F	DuMond Conservancy
Ana004	*Aotus nancymaae*	Nancy Ma's Owl Monkey	F	DuMond Conservancy
Ana005	*Aotus nancymaae*	Nancy Ma's Owl Monkey	F	DuMond Conservancy
Ana006	*Aotus nancymaae*	Nancy Ma's Owl Monkey	F	DuMond Conservancy
Av002	*Aotus vociferans*	Spix's or Noisy Owl Monkey	F	DuMond Conservancy
Hs001	*Homo sapiens*	Human	M	Self (P.Babb)
Hsy001	*Hylobates syndactylus*	Siamang	F	Coriell Institute (IPBIR: PR00969)
Ppa001	*Pan paniscus*	Bonobo (Pygmy Chimpanzee)	F	Coriell Institute (IPBIR: PR00092)
Ptr002	*Pan troglodytes*	Chimpanzee	F	Coriell Institute (IPBIR: PR00605)
Pab001	*Pongo pygmaeus abelii*	Sumatran Orangutan	M	Coriell Institute (IPBIR: PR00253)
Cge001	*Callitrhix geoffroyi*	White-fronted Marmoset	F	Coriell Institute (IPBIR: PR01094)
Papio001	*Papio Anubis*	Olive Baboon	M	Coriell Institute (IPBIR: PR00036)
Lla001	*Lagothrix lagotricha*	Woolly Monkey	F	Coriell Institute (NIA: NG05356)
Mamu001	*Macaca mulatta*	Rhesus Macaque	F	Coriell Institute (NIA: NG06249)
Sfu001	*Saguinus fuscicollis*	White-lipped Tamarin	F	Coriell Institute (NIA: NG05313)
Sss003	*Saimiri sciureus*	Squirrel Monkey	F	Coriell Institute (NIA: NG05311)
Hs002	*Homo sapiens*	Human	Unk	Promega Control DNA (pooled)

Note: Core study area is located in Formosa Province, Argentina (Lat. = 25º, 59.4 min South; Long. = 58º, 11.0 min West).
[a] = Control *A. azarai* sample used in this study.

(*Pithecia pithecia*) and squirrel (*Saimiri sciureus*) monkeys[7], which we also obtained from the Zoological Society of San Diego. Two human DNAs were also included as positive controls (Hs001: P. Babb, Hs002: Promega pooled DNA). Furthermore, we purchased comparative samples for 10 other primate species from the NIA and IPBIR cell line collections, curated by the Coriell Institute of Biomedical Research (Camden, NJ) (**Table 4.1**).

We also conducted searches of the publically available genomes accessible thorugh the University of California-Santa Cruz Genome Browser (UCSC GB) and the European Bioinformatics Institute's ENSEMBL browser for the extraction of fully annotated primate and rodent *AVPR1A* sequences (Kent et al., 2002; Karolchik et al., 2008). These searches yielded the entire A*VPR1A* locus for 11 primate and rodent species (**Table 4.2**). In total, we analyzed 25 *A. azarai* individuals for our intraspecific analyses, and 37 individuals (inclusive of AaaPLunk) representing 24 different primate and rodent species and subspecies for our interspecific analyses.

[7] Like owl monkeys, saki monkeys exhibit socially monogamous behavior, whereas squirrel monkeys are known to live in multimale-multifemale groups that vary in their tenure and complexity (Di Fiore and Rendell, 1994; Boinski, 1999; Rendell and Di Fiore, 2007).

Table 4.2. Comparative samples used in *AVPR1A* sequence analyses.

ID	Species	Common Name	Chromosomal Location of *AVPR1A*	Genome Browser/Portal
hg19	*Homo sapiens*	Human	12	UCSC Genome Browser
panTro3	*Pan troglodytes*	Chimpanzee	12	UCSC Genome Browser
gorGor3	*Gorilla gorilla gorilla*	Gorilla	12	UCSC Genome Browser
nomLeu1	*Nomascus leucogenys*	Gibbon	unknown	UCSC Genome Browser
ponAbe2	*Pongo pygmaeus abelii*	Orangutan	12	UCSC Genome Browser
rheMac2	*Macaca mulatta*	Rhesus Macaque	11	UCSC Genome Browser
calJac3	*Callithrix jacchus*	Marmoset	9 (partial)	UCSC Genome Browser
mm9	*Mus musculus*	Mouse	10	UCSC Genome Browser
rn4	*Rattus norvegicus*	Rat	7	UCSC Genome Browser
otoGar3	*Otolemur garnettii*	Bushbaby	unknown	Ensembl
micMur1	*Microcebus murinus*	Mouse Lemur	unknown	Ensembl

Sequencing

To investigate population-level variation at the *AVPR1A* locus, we targeted the 1.5 kb coding region of the *AVPR1A* gene and 10.5 kb of adjacent non-coding sequence (Murasawa et al., 1995; Thibonnier et al., 1996; Kent et al., 2002). This region encompassed both of the exons that encode the *AVPR1A* protein (1.5 kb), the 1 kb intronic region separating them, 1.9 kb of the 5' untranslated region (UTR), 1.0 kb of the 3' UTR, and 6.6 kb of the 5' flanking sequence upstream of the transcriptional start site (TSS) (**Figure 4.2**).

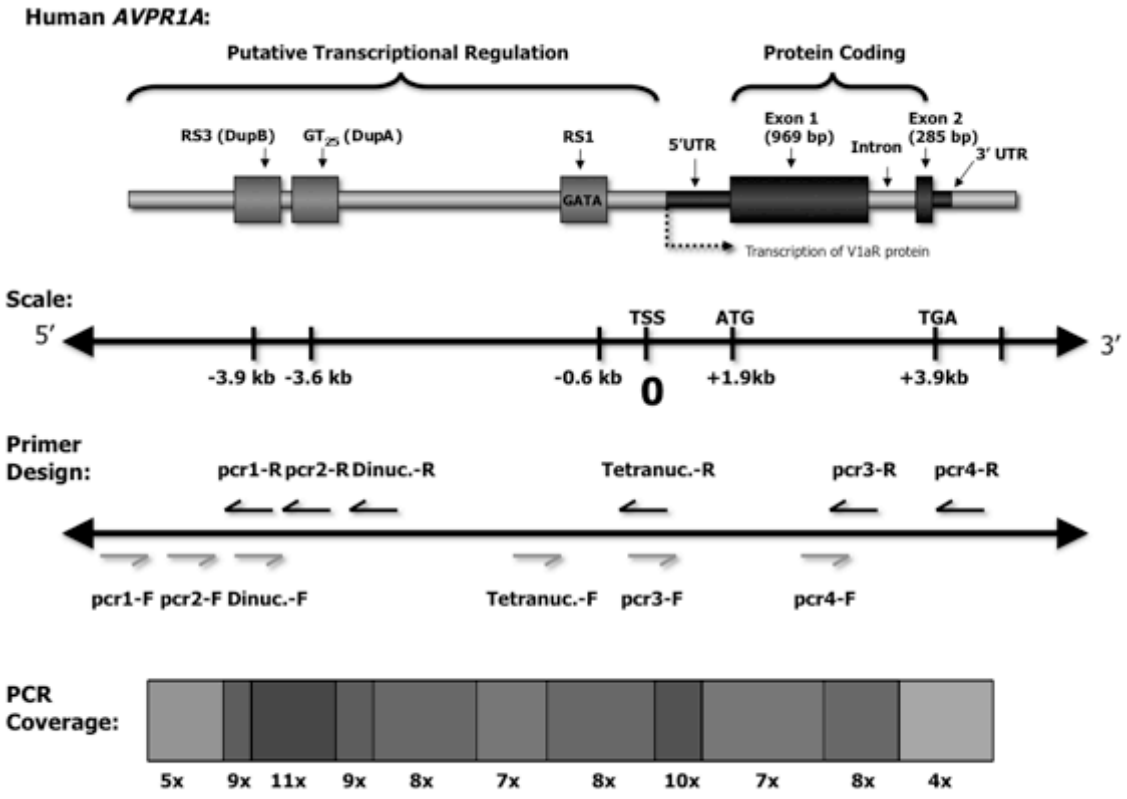


Figure 4.2. A schematic diagram of the *AVPR1A* gene in humans. Notable architectural landmarks of this gene are depicted on a scale beneath the figure (Thibonnier et al., 1996). The locations of the oligonucleotide primers used in this study for PCR amplification are shown in their approximate positions given the diagrammatic scale, as well as the coverage achieved by using different primer pairings for DNA amplification, sequencing, and fragment analysis.

We generated an interspecific *AVPR1A* contig from the sequence files retrieved from UCSC GB, and ENSEMBL using Sequencher, v4.9 (Gene Codes) and Geneious Pro v5.5.5 (Drummond et al., 2010). From the consensus sequence of this assembly, we designed 18 unique *AVPR1A* primers to bind in regions exhibiting high levels of sequence conservation across all taxa, using the oligo software programs NetPrimer (Premier BioSoft) and Primer3 (Rozen and Skaletsky, 2000). These primers amplified four overlapping fragments of ~1000 bp in length, with alternate primers situated within amplicons to facilitate sequencing (**Table 4.3**).

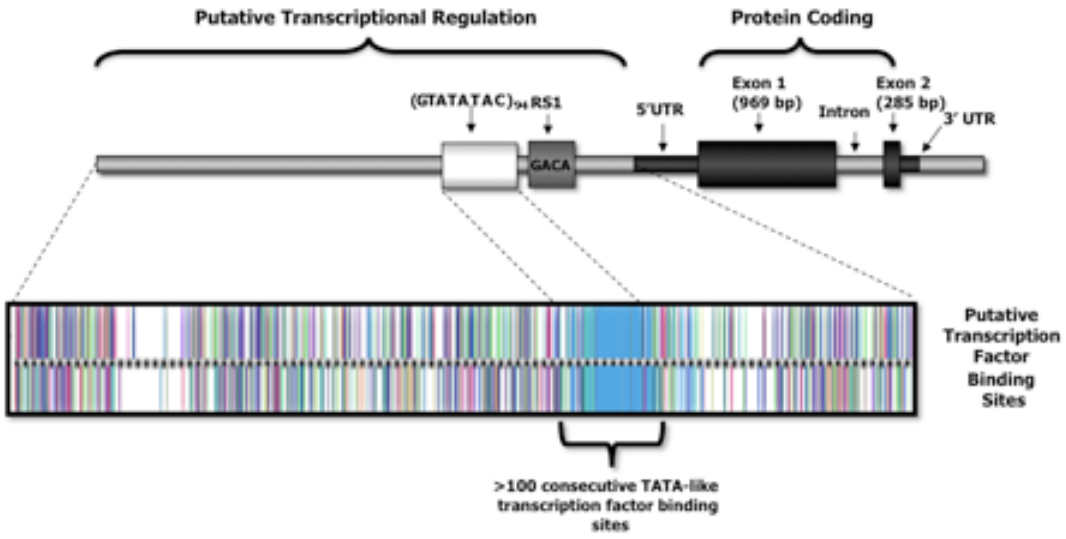
Table 4.3. Primer sequences for *AVPR1A* amplification and sequencing

Genetic Region	Primer Name	Strand	Oligonucleotide Sequence (5'-3')
5' Upstream Region (Long-range, all species)	AVPR1A-LR01-F[a]	Forward	TGT GAT TAA GGG GAA AAA ATA G
	AVPR1A-LR02-F[a]	Forward	AGT ATC CAC CTT GTA TCC TCT TG
	AVPR1A-LR03-F[a]	Forward	CCT TCT CTG GGC ACA CAG
	AVPR1A-LR04-F[a]	Forward	AAG TGG CTC AAC AGT CAA AG
	AVPR1A-LR05-F[a]	Forward	TCA GGG AGG GAT TTG GAA TG
	AVPR1A-LR01-R[a]	Reverse	CAC GGC GAT CTC CAG TTT G
5' Upstream Region (Upstream, GT_{25}/DupB)	V1aR-pcr1-F[a]	Forward	AGG CAC AGT GGC TCA TAC CT
	V1aR-pcr1-R[a]	Reverse	GCA AAA CTG CTG ACC ATG AA
5' Upstream Region (GT_{25}/DupB repeat)	V1aR-pcr2-F[a]	Forward	AAC CAT TTA AGT CCC TTC C
	V1aR-pcr2-R[a]	Reverse	GGT TTT TGG GTA TGC ATT GTG
5' Upstream Region (RS3: DupA repeat)	Dinucleotide-F [b,c]	Forward	GTA TTG CCA CAA ATA GAC CAA CG
	Dinucleotide-R [b,c]	Reverse	GTA AGG ATG ACA GGC GTT ACT G
5' Upstream Region (RS1: GATA repeat)	Tetranucleotide-F [d]	Forward	TAA TAC GAC TCA CTA TAG GG
	Tetranucleotide-R [d]	Reverse	CGC AAG CTT GGC ACT GCG TGC AGC TCT GCT CTG C
Coding Region (5'UTR, Exon 1, Intron)	V1aR-pcr3-F[a]	Forward	AGG ACA AAC ACC GAC GTA GG
	V1aR-pcr3-R[a]	Reverse	GCC TAC GTG ACC TGG ATG AC
Coding Region (Intron, Exon 2)	V1aR-pcr4-F[a]	Forward	CCG CAG TAC TTC GTC TTC TC
	V1aR-pcr4-R[a]	Reverse	TCT TCC AAG TCC ATC AAA TTC A
Coding Region (Exon 2, 3' UTR)	AVPR1A-Ex2-F[a]	Forward	GTA GCC ATT CCG TAA CTT CTG TG
	AVPR1A-Ex2-R1[a]	Reverse	GCT AGG GTG GTT ATA CTT TTT CC
	AVPR1A-Ex2-R2[a]	Reverse	TCT AAA TAC TAA GGC TGA TGA TGA G
	AVPR1A-Ex2-R3[a]	Reverse	CAC TTT ACA TCC TGA GTC ACT CTT TAG G

[a] = Current study
[b] = Thibonnier et al., 2000
[c] = Hammock and Young, 2005
[d] = Kim et al., 2002

Panel a.

Owl Monkey *AVPR1A*:



Panel b.

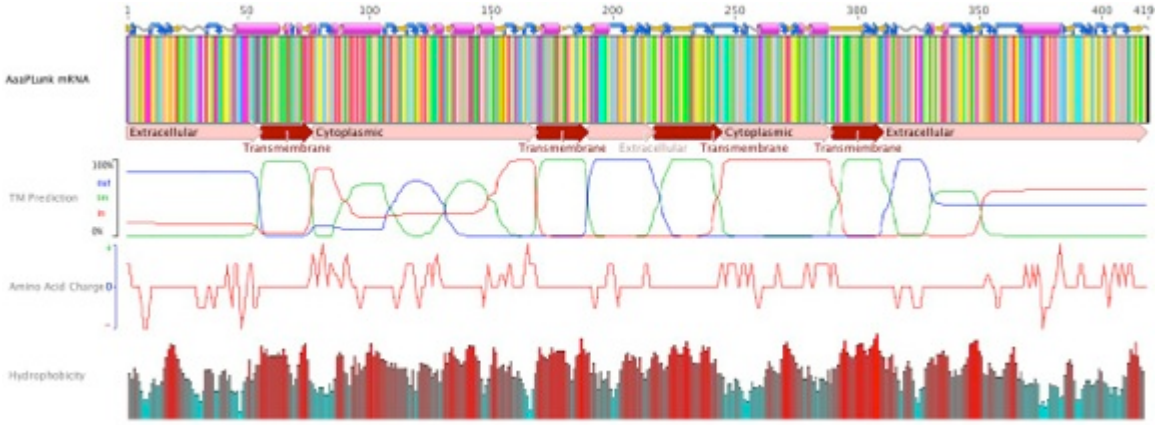


Figure 4.3, Panel a. The structure of the *AVPR1A* locus in *A. azarai.* The localization of putative transcription factor binding sites is identified by the boxed area, which is located upstream of the *AVPR1A* coding sequence. Over 100 consecutive TATA-like TFBS were identified in MatInspector (GenoMatix Software, GmbH). **Panel b.** The structure of the *AVPR1A* mRNA in *A. azarai*. Transmembrane predictions and hydrophobicity estimates are shown directly beneath the amino acid composition of the linearized mRNA sequence.

PCR cycling parameters were optimized for each primer pair using the Touchgene Gradient thermocycler (Techne), and all subsequent reactions were PCR amplified in GeneAmp 9700 thermocyclers (ABI). Amplified products were visualized on 1% TBE SeaKem agarose (Lonza) via gel electrophoresis.

Amplicons were purified and sequenced using previously described methods (Chapters 2 and 3). We assessed read quality for each sequence using Sequencing Analysis v5.4 (ABI) and aligned them using Sequencher and Geneious. This approach generated >10.5 kb of reliable *AVPR1A* genic sequence for each individual.

Sequence alignments

For each individual, we assembled full *AVPR1A* genic sequences by aligning the overlapping forward and reverse fragments using strict (>95%) agreement thresholds. To investigate variation present in different regions within the *AVPR1A* locus, we constructed separate sequence matrices to examine the nucleotide diversity in both the coding and non-coding regions of the gene for all individuals. These matrices isolated sequence variation within 5' UTR (379 bp in putative promoter), Exon 1 (563 bp), Intron (336 bp), Ex2 (287 bp), and 3' UTR (255 bp) regions in both our intra- and interspecific analyses. In addition, we obtained and assembled coding sequences representing the entire mRNA of the *AVPR1A* gene (1257 bp / 418 amino acids + TGA) in all *A. azarai* individuals, and compared the structure of the owl monkey V1a protein to the complete genome sequences from UCSC GB and ENSEMBL.

To ensure the amplification and identification of repeat motifs in the *AVPR1A* 5' upstream region known to exist in other mammals, we designed several additional PCR and sequencing protocols. These involved alternative primer combinations (e.g., using pcr1-F with pcr2-R), touchdown PCR (TD-PCR) cycling conditions with increased extension times, and direct gel excision and purification techniques (Don et al., 1991; Korbie and Mattick, 2008). Upon excision, specific amplified PCR fragments were

purified and sequenced as described above.

Using this collection of sequences and STR alleles, we searched for any polymorphic mutations and putative transcription factor binding sites (TFBS) in the upstream promoter the *AVPR1A* gene in *A. azarai* relative to other primate species. This study was conducted using the MatInspector software package (GenoMatix Software, GmbH), with parameters being set for transcription factors present in mammalian neuronal cellular tissues as derived from human and mouse consensus databases. This work was later confirmed using the transcription factor prediction software plugin for Geneious v5.5.5 (Drummond et al., 2010).

Population genetic analysis

Summary statistics, including pairwise identity, mean number of pairwise differences (π), GC content, and haplotype (h) diversity, were calculated for all *AVPR1A* sequence matrices using programs available in Arlequin v3.5 (Excoffier et al., 2005) and DnaSP v4.50 (Rozas et al., 2003). We also calculated the neutrality indexes of Tajima's *D* (Tajima, 1989a, b) and Fu's F_s (Fu, 1997) to explore the demographic and evolutionary forces acting on the coding region of *AVPR1A* in our study population. Transversions (TV) were weighted higher than transitions (TI) (10:1 final ratio) in all statistical calculations to account for the molecular probability for either occurring across evolutionary time (Bandelt et al., 1995, 1999, 2000, 2002).

Phylogenetic analyses: Data alignment for phylogenetic analyses

For phylogenetic analyses of *AVPR1A* sequences from platyrrhine and other primate species, we restricted the number of *A. azarai* sequences used in the analyses to AaaPLunk, which possessed the most frequent haplotype (identified in our study population) for each of the delineated genic regions (5'UTR, Ex1, Intron, Ex2, 3'UTR). This decision was made to avoid biasing the range of sequence variation toward the species *A. azarai*.

For each of the five predefined genic regions, we aligned the AaaPLunk *A. azarai* sequence with the 11 primate and rodent *AVPR1A* sequences obtained from the UCSC GB and ENSEMBL, as well as with the appropriate sequences derived from our 25 comparative samples. The resulting interspecific matrices, each consisting of 37 sequences representing 24 species and subspecies, were pruned to avoid the ambiguous characters (*N*) found in some of the public genomes, and were carefully edited to account for insertion-deletion polymorphisms. The comparative dimensions of each matrix, and the number of useable sites for each genetic region can be found in **Tables 4.4** and **4.5**.

The resulting matrices were annotated in Geneious (Drummond et al., 2010), then imported into MacClade (Maddison and Maddison, 2003) for character and taxonomic organization, weight designation, and appropriate command-block formatting for phylogenetic tree calculations in PAUP* v4.0b10 (Swofford, 2002), PAML/codeML (Yang, 2007), and MrBayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The output files from these analyses were imported back into MacClade as well as FigTrees v1.2.3 (A. Rambaut, http://beast.bio.ed.ac.uk/FigTree) for tree visualization and phylogram generation.

Phylogenetic analyses: Network analysis

We generated multistate median joining (MJ) and binary reduced median (RM) networks from all *A. azarai AVPR1A* sequences using Network v4.5.02 (Bandelt et al., 1995, 1999, 2000, 2008). Both types of networks were constructed to assess the consistency of the observed relationships among *AVPR1A* haplotypes, and test the effect of removing terminal reticulations from the phylogenetic branches (Bandelt et al., 1999). We again employed a 10:1 TV:TI weighting scheme (Bandelt et al., 1995, 1999, 2000) with slight modifications for the construction of networks in order to reduce the phylogenetic reticulations caused by homoplasies and hypervariable characters.

We similarly produced MJ and RM networks for the sequences of *AVPR1A* from comparative primate and rodent taxa to explore the degree of nucleotide variation in the gene across broader evolutionary distances. In this case, we used an 850 bp combination of sequences representing Exon 1 (positions 1-563 of 969 bp) and Exon 2 (positions 1-287 of 287 bp) that we were able to reliably generate or download for all species.

Phylogenetic analyses: Model selection

To select the most appropriate model for our phylogenetic analyses of *AVPR1A* Exon 1 (563 bp), we ran the program jModelTest v0.1.1 (Guindon and Gascuel, 2003; Felsenstein, 2005; Posada, 2008) using 11 substitutions patterns to survey 88 models of nucleotide substitution (+F base frequencies, rate variation of +I and +G with nCat=4). We implemented the modified AICc setting due to the small size of comparative nucleotide characters (563 bp), and conducted parallel searches using BIC and DT parameters. The base tree for our likelihood calculations was optimized for maximum

likelihood (ML) phylogenetic analysis.

Phylogenetic analyses: ML and BI analyses

We conducted ML analysis in PAUP* v4.0b10 (Swofford, 2002) to estimate the most likely evolutionary tree based on the alignments of *AVPR1A* nucleotides and amino acid codons and using the nucleotide model specified by our jModelTest runs. For ML analysis, we estimated bootstrap values based on a set of 10,000 replicates.

To find the phylogenetic tree or set of trees that maximizes the probability of obtaining our data given a specified model of evolution, we undertook BI analysis with the software program MrBayes v3.1.2 (Huelsenbeck and Ronquist, 2001; Huelsenbeck et al., 2001; Ronquist and Huelsenbeck, 2003). In this analysis, we used the ML function and employed six substitution types (nst=4) with the BIC-specific nucleotide substitution model as suggested by jModelTest. The MCMC search was run with four chains for 100,000 generations, with trees sampled every 100 generations. Using the average standard deviation in split frequencies among the four chains, we were able to assess the level of convergence (chain deviation <0.05) denoting an acceptable level of post-convergence tree likelihoods that influence the accuracy of our consensus Bayesian tree. The first 1000 trees were discarded as "burn-in" to remove extraneous pre-convergence probability values that can skew the values of a given consensus Bayesian tree (Huselsenbeck et al., 2001; Altekar et al., 2004).

We also employed MP and ME/NJ tree-building methods to identify any incongruence between the different algorithms. We estimated bootstrap values for MP and ME/NJ trees based on a set of 10,000 replicates.

Phylogenetic analyses: Analysis of adaptive evolution

To identify the level of *AVPR1A* codon variation in our mRNA matrix (12 sequences by 1276 bp, gapped), we examined the non-synonymous-to-synonymous amino acid substitution ratio (d_N/d_S) in the *AVPR1A* exons 1 and 2, which represent all codons in *A. azarai* versus the 11 complete genome sequences. Applying the ML tree (identified through likelihood analysis in PAUP*) to the codeML program within PAML (Yang, 2007), we calculated the relative rates of amino acid changes along the different phylogenetic branches. This analysis was undertaken to detect any possible signatures of adaptive evolution and directional selection, such as an excess of non-synonymous mutations along a particular taxonomic lineage. We ran codeML twice to produce two phylogenetic models (M0 and M1), which were then subjected to a likelihood ratio test (LRT) to assess the significance of estimating individual d_N/d_S ratios (variable) versus estimating one ratio (homogeneous) for the entire tree.

Phylogenetic analyses: Bayesian coalescent estimation

To check for the presence of *AVPR1A* mutation rate variation among the 24 species and subspecies, and to assess the impact that a relaxed molecular rate of evolution would have on age estimations for hypothetical ancestral *AVPR1A* gene sequences, we analyzed our coding region matrix in the software program BEAST v1.5.3 (Drummond et al., 2006; Drummond and Rambaut, 2007). This program takes into account the potential errors associated with a fixed mutation rate through Bayesian MCMC calculation of aligned molecular sequences (Ho et al., 2005), and generates a range of coalescent events based on prior knowledge of calibrated paleontological events (Drummond et al., 2002;

Drummond and Rambaut, 2007). We used the companion software program BEAUti v1.5.3 (Drummond et al., 2006; Drummond and Rambaut, 2007) to apply time points of 31 mya (SD 0.5 mya) for the platyrrhine-catarrhine time to most common ancestor (TMRCA), and 20 mya (SD 1.0 mya) for the TMRCA of *Aotus*, *Saimiri,* and *Callithrix* (applied priors outlined in Babb et al., 2011).

By implementing the Yule speciation process parameters in BEAST, we specified the abovementioned time points as log normally distributed priors applied to the appropriate taxon designations. We used the relaxed-clock lognormal molecular clock and ran the software using the nucleotide model specified by our jModelTest runs. BEAST was run for 2,000,000 generations, echoing on-screen every 10,000 and logging every 200.

The Bayesian coalescent results generated in BEAST were analyzed with the companion software in TRACER v1.5 (Rambaut and Drummond, 2007) to view the distribution of coalescent time points for our prior specified monophyletic taxonomic groups. The 10,001 trees retained from the BEAST run were summarized in TreeAnnotator v1.5.3, and displayed in the program FigTrees v1.2.3.

Genomic comparisons

To investigate transcriptional control of expression of *AVPR1A* in primates, we compared our aligned sequences and repeat motifs to those in the marmoset genome (*Callithrix jacchus*) on UCSC Genome Browser (Kent et al., 2002; Karolchik et al., 2008). We noted that the region overlapped several architectural contigs of the previous two versions of marmoset genome (calJac1, calJac2), denoting regions located on

different chromosomes. The recent version, calJac3, now denotes the chromosome as number 9, but significant gaps still exist. Due to these architectural gaps, we subsequently broadened our comparison to encompass additional genomes (panTro3, calJac3, gorGor3, ponAbe2, nomLeu1), using the *Homo sapiens* (hg19) genome as a reference track. In addition, we applied the most recent annotation of these regions using the Database of Genomic Variants (DGV) (Iafrate et al., 2004) analytical track on UCSC GB.

4.3 Results

AVPR1A sequence diversity in Aotus

Our analysis of 25 owl monkeys reveals a low level of diversity in the 1257 bp coding region of the *AVPR1A* gene. Among these individuals, we observed a complete lack of genetic variation (ts, tv, and insertion-deletion substitutions) within the *AVPR1A* coding region matrices: specifically, Exon 1, Exon 2, and the intervening Intron. Along the same lines, the pairwise identity for all *A. azarai* sequences was nearly 100% identical in each of the 5 genic regional data matrices.

However, we did observe five single nucleotide polymorphisms (SNPs) within the 5' and 3' UTRs, that occurred at low to medium frequencies (6% - 38%) within the 25 *A. azarai* individuals. As a result, these SNPs defined five distinct haplotypes, defined by this variable non-coding UTR sequence data (**Table 4.4**), and confirmed by the results of our intraspecific network analysis. Overall, these findings coincide with low levels of intraspecific sequence diversity observed in other studies of the *AVPR1A* coding region in mammalian taxa (Fink et al., 2006, 2007; Phelps et al., 2009; Turner et al., 2010).

Table 4.4. Summary statistics for *AVPR1A* sequences in 25 Azara's owl monkeys (*A. azarai*).

	5' UTR 379 bp	Exon 1 563 bp	Intron 336 bp	Exon 2 287 bp	3' UTR 255 bp
Sequence diversity:					
Pairwise identity of alignment (%)	99.7%	100%	100%	100%	99.97%
GC content (% non-gap)	64.9%	65.4%	25%	41.8%	32.2%
Summary statistics:					
Usable loci (<5% missing data)	379	563	336	287	255
Identical sites	375 (98.9%)	563 (100%)	336 (100%)	287 (100%)	254 (99.6%)
Polymorphic sites	4	0	0	0	1
Single nucleotide polymorphisms (SNPs) [freq.]	024C>T [10%] 222T>C [26%] 225A>G [38%] 228G>T [6%]	0	0	0	013T>C [4%]
Haplotypic diversity:					
# Haplotypes	5	1	1	1	2
# Singletons	1	0	0	0	0

Interestingly, our analysis of the 5' upstream region of the *AVPR1A* gene of *A. azarai* revealed a novel STR pattern that distinguished owl monkeys, both in sequence and repeat number, from other primate taxa (**Figure 4.3**). *A. azarai* individuals possessed a GACA repeat pattern similar to that of the marmoset at the RS1 location (**Figure 4.1**). However, no GT_{25} or RS3 sequence motifs were observed in any of the owl monkey individuals. On the other hand, we did encounter a palindromic $GTATATAC_{94}$ repeat roughly 4000 nucleotides upstream of the *A. azarai AVPR1A* start site. Unfortunately, this region was denoted by a run of 712 "N[8]"s in the annotated *C. jacchus* genomic sequence on the UCSC GB, thereby preventing us from confirming the presence of this repeat in marmosets. We were also not able to obtain the sequence containing the

[8] The standard International Union of Pure and Applied Chemistry (IUPAC) naming convention for genetic nucleotid bases characterizes unknown and/or ambiguous base calls with the letter "N".

GTATATAC repeat in *Lagothrix*, *Callitrhix* (*geoffroyii*), *Saguinus*, *Pithecia*, or *Saimiri* to check for its presence in other platyrrhines.

Furthermore, we observed surprisingly little variation in this 5' upstream region in the owl monkey DNAs. More specifically, we observed 94 repeats of the GTATATAC in all 25 *A. azarai* individuals (**Figure 4.3**). The sequence of this entire repeat appeared to be fixed, even in individuals from widely disparate mtDNA lineages, or, in the case of four zoo animals, from individuals originating over 250 km away. This lack of STR variation in *A. azarai* individuals markedly differs from the comparable *AVPR1A* 5' upstream variation seen in monogamous voles with regard to the RS1 repeat, as well as that observed in humans and chimps, which vary at the RS3 microsatellite (**Figure 4.1**).

AVPR1A sequence diversity among platyrrhines

In our interspecific comparison of sequence diversity in 37 primate and rodent samples across five regions of the *AVPR1A* locus, we again saw lower levels of diversity in coding region sequences (nucleotide diversity: 0.33-0.24) compared to UTR/regulatory and intronic sequences (0.49-0.56). Nevertheless, we were still surprised by the amount of variation present, including the substantial numbers of insertion deletions present in the 5' and 3' UTRs (**Table 4.5**). For example, in the 410 bp matrix (gapped) representing the 5' region immediately upstream of the *AVPR1A* methionine start site (ATG), only 45 sites are identical (11% site identity) despite a 77.6% overall pairwise identity of the full alignment. The bulk of this diversity comes from clade-specific (e.g., platyrrhine-only, strepsirhine-only, hominoid-only, etc.) insertions or deletions of large strings of

nucleotides, and highlights the molecular instability of this non-coding region across primate and rodent genomes.

Table 4.5. Summary statistics for *AVPR1A* sequences in 37 primate and rodent samples (24 species and subspecies)

	5' UTR 410 bp*	Exon 1 563 bp	Intron 416 bp*	Exon 2 287 bp	3' UTR 283 bp*
Sequence diversity:					
Pairwise identity of alignment (%)	77.6%	92.6%	82.2%	93.3%	85.9%
GC content (% non-gap)	46.3%	65.9%	22.3%	41.9%	32.8%
Nucleotide diversity:					
Nucleotide diversity (avg. over loci)	0.51 +/- 0.25	0.33 +/- 0.16	0.56 +/- 0.27	0.24 +/- 0.12	0.49 +/- 0.24
Summary statistics:					
Usable loci (<5% missing data)	410	563	416	287	280
Identical sites	45 (11%)	350 (62.2%)	77 (18.5%)	176 (61.3%)	59 (20.8%)
Polymorphic sites	365	213	339	111	221
Sites with transitions (ts)	110	139	133	84	116
Sites with transversions (tv)	125	102	138	42	103
Insertion/deletions (indels)	290	24	192	0	69
Selective neutrality:					
Mean # pairwise differences (π)	173.05	180.78	208.50	69.45	128.62
Sites with substitutions (S)	181	200	220	111	221
Tajima's *D* (1000 simul.)	11.21	10.41	11.11	6.16	7.75
P (*D* simul < *D* obs)	0.03	0.02	0.03	0.001	0.008
Fu's F_S (1000 simul.)	-2.98	-3.38	-2.64	-8.25	-4.50
P (sim_F_S ≤ obs_F_S)	0	0	0	0	0

* = Indicates width of nucleotide matrix including alignment gaps (to account for insertion/deletion events).

Furthermore, each platyrrhine species possessed lineage-specific coding region nucleotide changes that distinguished it from the other genera. In particular, the mouse lemur (*Microcebus murinus*) and bushbaby (*Otolemur garnettii*) both have codons 17-22 completely deleted from their Exon 1 sequences. These codons are present in both the mouse (*Mus musculus*) and the rat (*Rattus norvegicus*), indicating that the deletion is a shared derived occurrence in these two primate species (**Table 4.6**).

Table 4.6. Summary statistics for complete *AVPR1A* mRNA in 12 primate and rodent species

	Spliced *AVPR1A* mRNA molecule 1276 bp*
Sequence diversity:	
Pairwise identity of alignment (%)	88.6%
GC content (% non-gap)	57.7%
Nucleotide diversity:	
Nucleotide diversity (avg. over loci)	0.49 +/- 0.25
Summary statistics:	
Usable loci (<5% missing data)	1276
Identical sites	823 (64.5%)
Polymorphic sites	453
Sites with transitions (ts)	292
Sites with transversions (tv)	198
Insertion/deletions (indels)	45
Selective neutrality:	
Mean # pairwise differences (π)	612.11
Sites with substitutions (S)	418
Tajima's *D* (1000 simul.)	16.15
P (*D* simul < *D* obs)	0.03
Fu's F_S (1000 simul.)	2.20
P (sim_F_S ≤ obs_F_S)	0
Amino acid diversity:	
Mean molecular weight (kDa)	46.819
Mean isoelectric point	9.69

* = Indicates width of nucleotide matrix including alignment gaps (to account for insertion/deletion events). The minimum sequence length was 1233 bp (411 amino acids) belonging to otoGar1 and micMur1, whereas the maximum sequence length was 1275 bp (425 amino acids) was observed in mm9 and rn4.

To investigate the selective neutrality of *AVPR1A* in primates and rodents, we calculated Fu's F_S and Tajima's *D* statistics for the 37 sequences in the five different regions of the locus. This analysis revealed Tajima's *D* values ranging from 11.21 (*P*<0.03) in the 5' UTR to 6.16 (*P*<0.001) in Exon 2. Fu's F_S values ranged from -2.98 (*NS*) in the 5' UTR down to -8.50 (*NS*) in Exon 2 (**Table 4.4**). The high positive value of Tajima's *D* indicated that an excess of intermediate frequency alleles was present in our data set, and suggested the possibility of balancing selection of heterozygosity (in the form of nucleotide diversity) at the *AVPR1A* locus across this phylogenetic sampling of

primate and rodent genomes. The negative value of Fu's F_S generally confirmed this interpretation.

Moving further upstream of the 5' UTR, all of the platyrrhine taxa we examined possessed a GACA microsatellite repeat sequence motif. This pattern distinguishes the platyrrhines from their catarrhine and hominoid relatives, all of which possess a GATA motif in same RS1 location (Donaldson et al., 2008; Rosso et al., 2008). The distinction of RS1 region in each of these platyrrhine lineages, all of which are assumed to have derived from a common ancestor that colonized the South American continent over 30 million years ago (Poux et al., 2006; Tejedor et al., 2006; Hodgson et al., 2009), may point to the existence of a specific New World motif of the *AVPR1A* gene promoter.

Phylogenetic analysis of AVPR1A: Network analysis

The network of *A. azarai AVPR1A* coding sequences revealed one distinct clade that consisted of a single central haplotype, which occurred at high frequencies in the population, and zero derivative haplotypes extending from it (**Figure 4.4a**). When we examined the coding sequences from a greater number of owl monkeys within our population (n=106), the number of haplotypes only grew by two, and each of those represented a singleton individual possessing a unique SNP. This pattern reflects noteworthy sequence conservation, and could be indicative of purifying selection acting upon the specific transcriptional code generated by *AVPR1A* locus within owl monkeys.

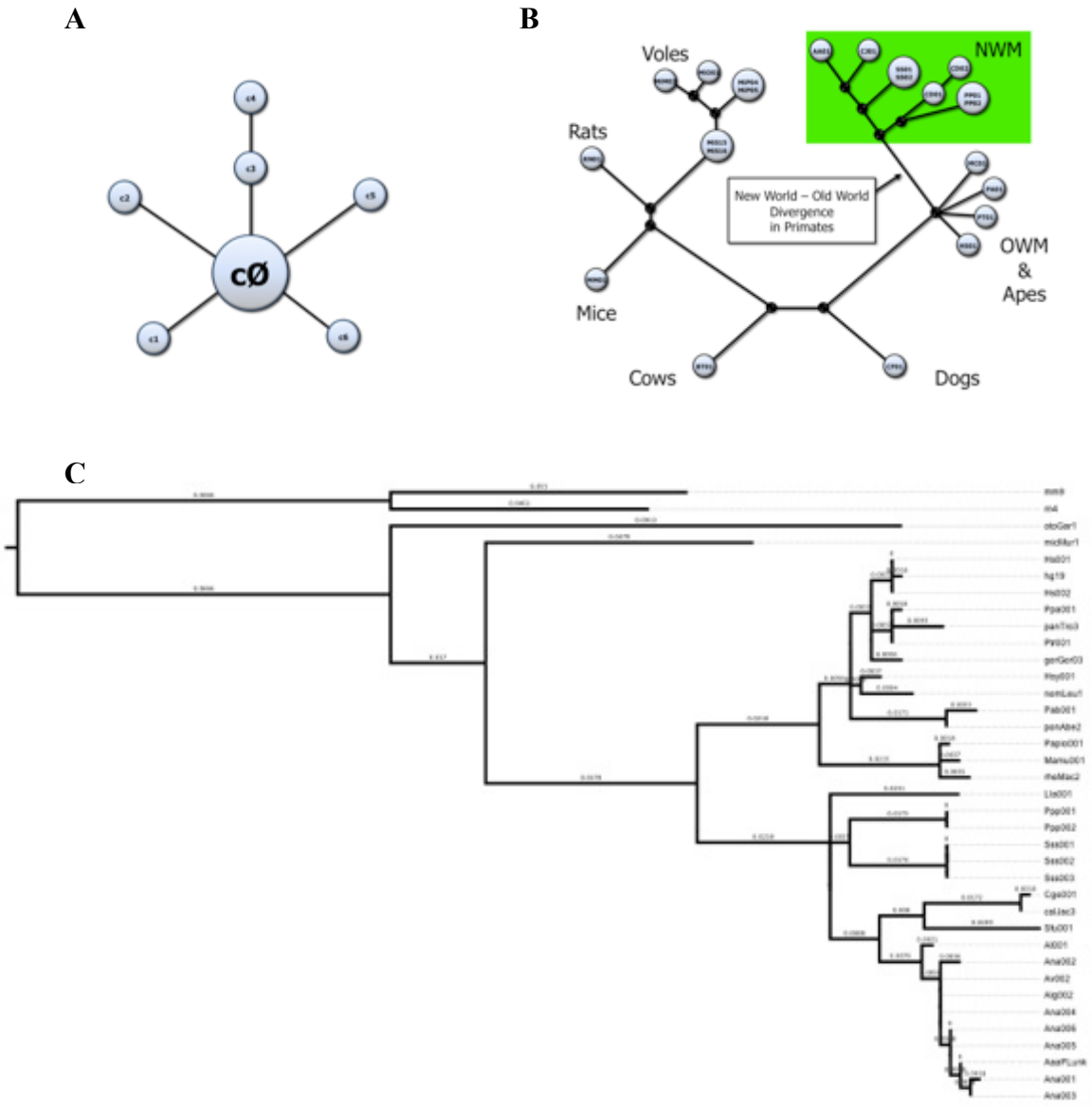


Figure 4.4. Panel a. A median joining (MJ) network of *AVPR1A* coding region sequences from *A. azarai* individuals. Node size is relative to the number of individuals that share a particular sequence. **Panel b.** A MJ network of *AVPR1A* coding region sequences from 17 mammalian and primate species. The branches leading to the different clades are proportional to their actual mutation distances, but have been shortened for the full representation of *AVPR1A* diversity in this network. As before, node size is relative to the number of individuals that share a particular sequence motif. **Panel c.** A ML phylogram of *AVPR1A* coding region sequences obtained from 37 primate and rodent species.

We also created networks for the comparison of *AVPR1A* coding sequences from primate and rodent species. In these networks, the *Aotus* sequence was separated from

other platyrrhine taxa (**Figure 4.4b**). Among platyrrhine taxa, *Lagothrix* and *Pithecia* were clearly distinguished from each other, and both were quite distant from *Saimiri*, whearas *Aotus* showed genetic affinities with *Callithrix* and *Saguinus*. In addition, a dramatic number of coding region changes separated the New World primates from the Old World monkeys and apes, with the latter showing considerably less mutational substructure and shorter internal branch lengths than the platyrrhines.

Phylogenetic analysis of AVPR1A: Phylogenetic model selection

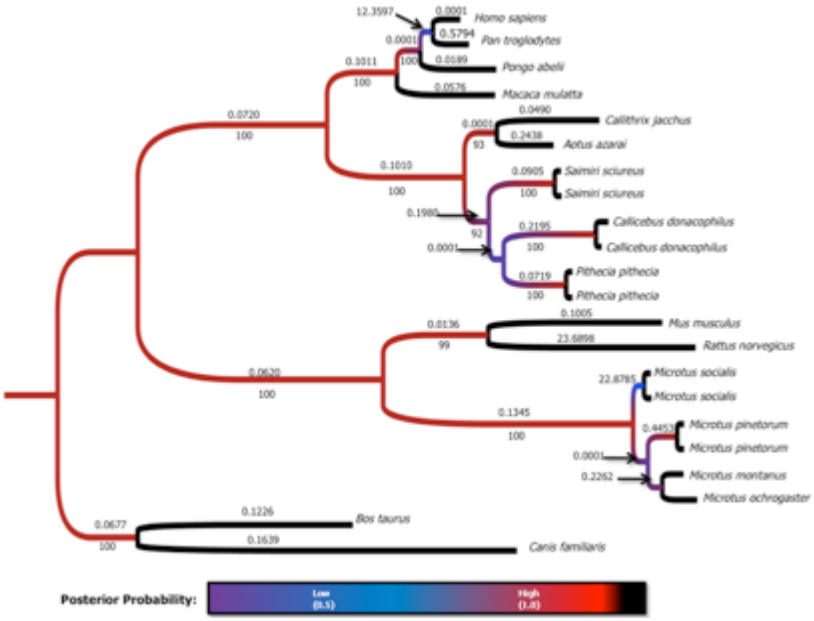
All three searches (AICc, BIC, DT) that we ran in jModeltest selected the HKY model (Hasegawa et al., 1985) for the phylogenetic model. In addition, all three search formats gave the HKY+G model a likelihood score (-ln *L*) of 2529.93. Furthermore, each selection processes specifically suggested the use of the HKY model plus gamma (+G) distribution for use in our phylogenetic calculations.

Phylogenetic analysis of AVPR1A: ML and BI analyses

Using the ML phylogenetic algorithm available in PAUP*, we investigated taxonomic relationships of 24 primate and rodent taxa based on nucleotide variation in the first 563 nucleotides of *AVPR1A* Exon 1. In agreement with the network analysis, the ML tree (**Figure 4.5a**) exhibited a similar breakdown of known taxonomic relationships of species based on genetic data (Hodgson et al., 2009), although the exact phylogenetic relationships that exist between platyrrhine genera is still an issue of considerable debate. Nevertheless, the majority of phylogenetic relationships among these species remained consistent across the different analyses we performed, with the exception of *Saimiri*,

Callithrix, and *Saguinus*, whose placement within the platyrrhines was dependent on the algorithm employed, mimicking the discrepancies noted by previous phylogenetic studies of platyrrhines (Opazo et al., 2006; Poux et al., 2006; Hodgson et al., 2009; Wildman et al., 2009).

Panel a.



Panel b.

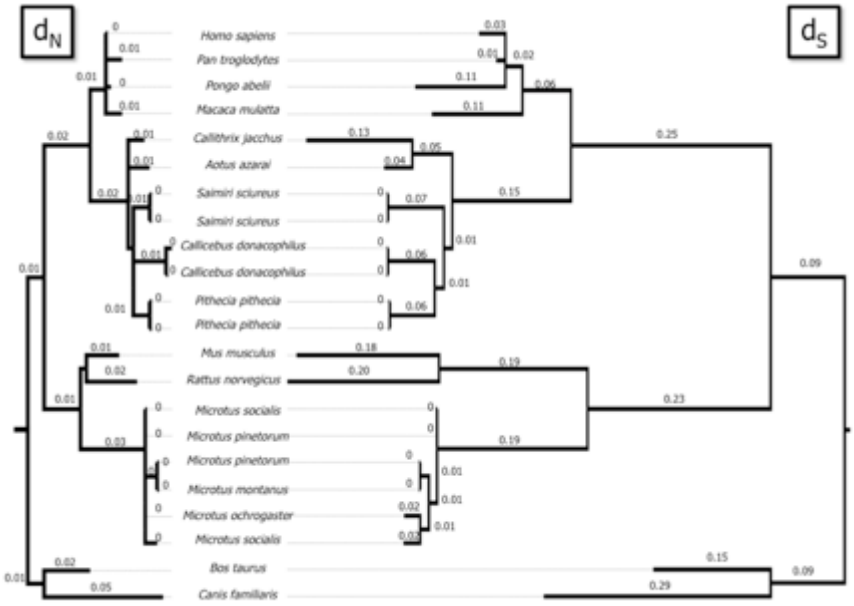


Figure 4.5. Panel a. A consensus phylogram representing the agreement of our ML and BI phylogenetic arrangements of 22 *AVPR1A* coding region sequences representing 17 different mammalian species. Branch lengths display the relative number of mutational differences detected for each different taxonomic group. The d_N/d_S values are shown above each branch, whereas ML bootstrap values (of 10,000 replicates) appear below them. For the BI analysis, two MCMC chains were run in parallel across 500,000 generations, with trees sampled every 100 generations. A total of 5,001 trees were retained and sampled for a 50% majority consensus tree, after a 1,000-tree burn-in. Posterior probabilities of the phylogenetic arrangement are displayed as a color gradient across taxa. **Panel b.** Oppositional phylogram representing the different phylogenetic patterns and branch lengths based on *AVPR1A* coding sequence variation and exhibited by d_N and d_S trees.

Phylogenetic analysis of AVPR1A: Analysis of adaptive evolution

Using the ML tree calculations, we predicted the relative rates of *AVPR1A* amino acid changes along the different phylogenetic branches in disparate mammalian taxa to detect any signatures of adaptive evolution. The resulting d_N/d_S ratio values for the majority of the branches were quite varied, likely reflecting processes of balancing selection at the codon level across mammals (Yang, 2007) (**Figure 4.5b**). Alternatively this diversity could have been shaped with more stochastic paleo-demographic or genetic drift events throughout mammalian history. We noted stronger positive selection signals ($d_N>d_S$) on branches leading to the taxonomic outgroups, with these being further heightened when comparing the pure d_N tree to the pure d_S arrangement. The LRT indicated that the M1 (variable) model was more likely (ln L = -2343.8922) than M0 (homogenous) model (-ln L = -2360.8002), although this result was non-significant (P>0.7796). Therefore, we decided that individual branch values would more accurately describe our data with regard to the adaptive evolution of *AVPR1A* and directional selection in primates.

Phylogenetic analysis of AVPR1A: Bayesian coalescent estimation

The BEAST analysis produced a highly resolved phylogenetic chronogram of primate and mammalian species with *AVPR1A* coalescent time points for each branch of the tree. Its topology presented a taxonomic arrangement of these species that was very similar to that produced by both the ML and median network analyses (**Figure 4.6**). In addition, through this Bayesian analysis, we were able to visualize the changes in mutation rate that occurred along different evolutionary branches, as well as generate age

estimations at each phylogenetic node. The TMRCA estimations and their associated error ranges were consistent with those in other recent studies of molecular data (Hodgson et al. 2009). These conditions strengthened our confidence in the evolutionary scenario for *AVPR1A* among the *A. azarai* and platyrrhine primates that we propose below.

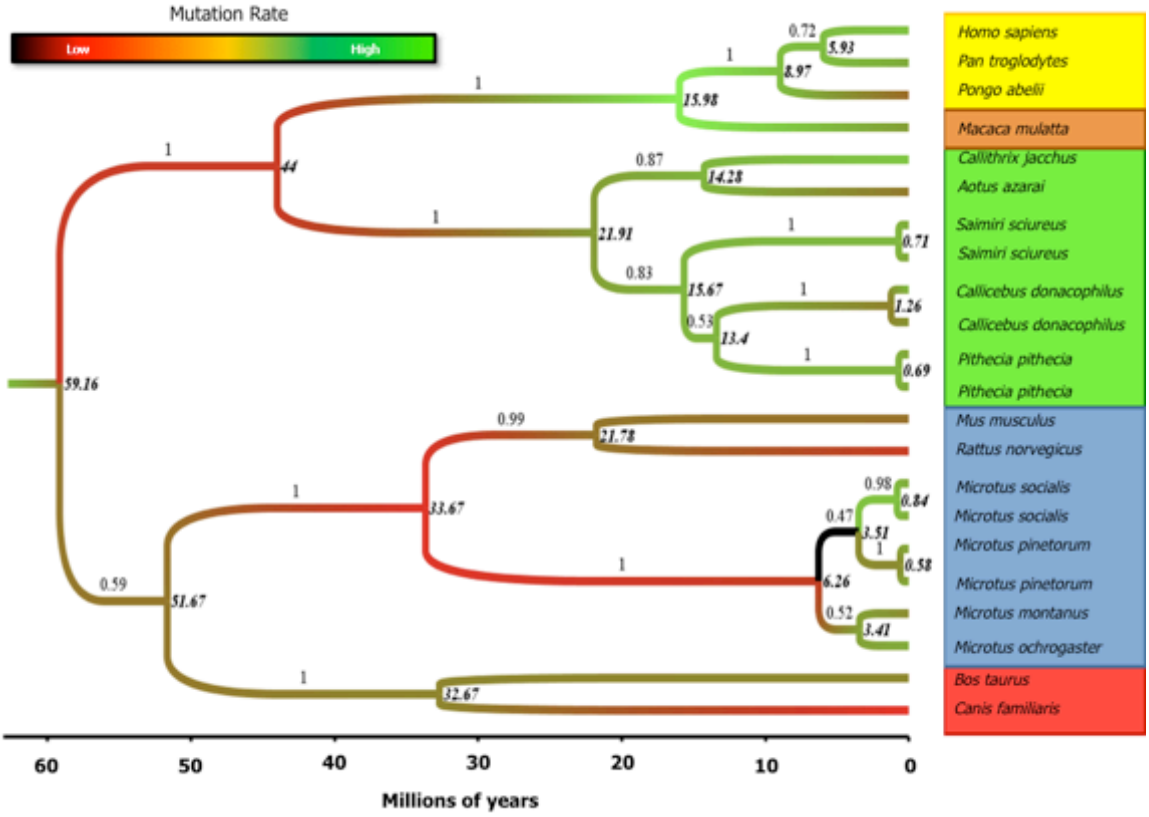


Figure 4.6. A chronogram depicting the Bayesian coalescence age estimation of *AVPR1A* coding region sequences, as calculated in BEAST v1.5.3. Age estimates for each node are italicized and positioned to the right of them. Posterior probabilities are indicated above each branch. Mutation rate of *AVPR1A* along the different evolutionary lineages is displayed as a red-green color gradient across taxa. The different taxa compared here are displayed on colored backgrounds corresponding to their phylogenetic placement. *Yellow* large bodied apes, *gold* small bodied apes, *orange* Old World monkeys, *green* New World monkeys, *blue* rodents, *red* other Eutherian mammals (i.e., neither rodent nor primate).

4.4 Discussion

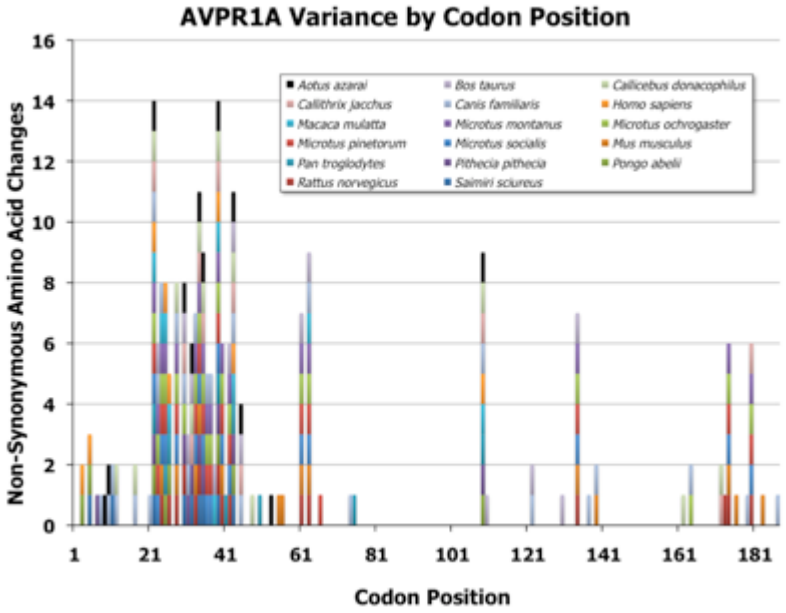
AVPR1A evolution in A. azarai

The individuals from our owl monkey population exhibit very little coding region

diversity at the *AVPR1A* gene. In addition, the repetitive motifs present in the 5' upstream sequence of *AVPR1A* are conserved in all individuals that we examined. In fact, *A. azarai* individuals exhibit complete uniformity in the putative repeat sequences (both RS1 [GACA] and $GTATATAC_{94}$) present in the putative *AVPR1A* regulatory region. The 5' and 3' UTRs directly adjacent to Exons 1 and 2, respectively, presented the only notable polymorphism observed in different *A. azarai* individuals. Thus, there appears to be a lineage-specific pattern of sequence conservation at this locus in this species.

One scenario that would explain these data is that varying degrees of selection are taking place at the repeat motifs and at the coding region. Selection occurring at the regulatory region could serve the purpose of altering functional gene expression, whereas conservation at the coding region could preserve particularly adaptive protein structures. The coding region mutations observed in *Aotus* (**Figure 4.7a**), relative to *AVPR1A* coding sequences identified in other taxa, imply that such structural alterations have taken place at the ligand and G-protein binding domains in the mature *AVPR1A* protein (**Figure 4.7b**). These non-synonymous changes could affect the functional activity of AVP receptor neurons, and ultimately influence the manifestation of the sociobehavioral phenotype of long-term partner preference.

Panel a.



Panel b.

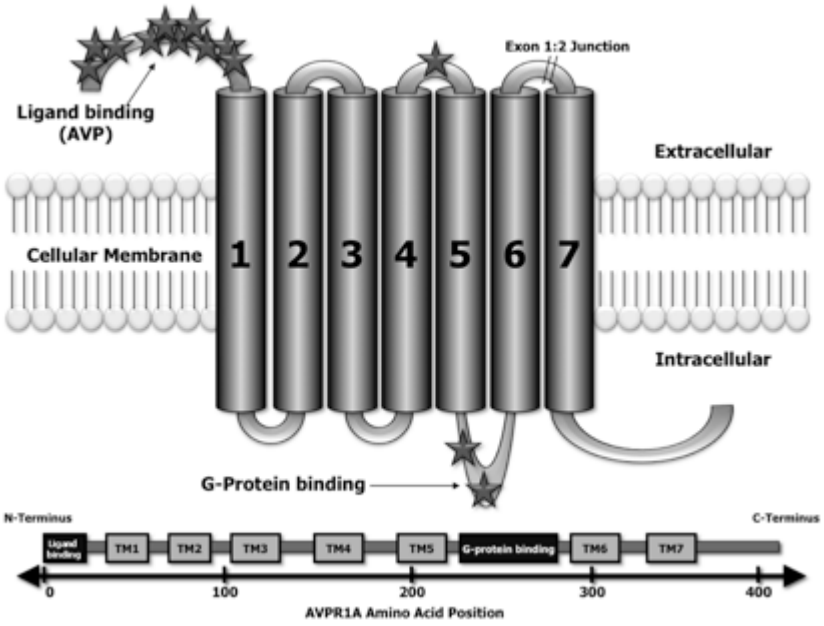


Figure 4.7. Panel a. A stacked histogram displaying the number of non-synonymous amino acid changes of the *AVPR1A* coding sequence in 17 different species, relative to a translated consensus DNA sequence (assessed by plurality, [50% majority rule]). Black bars represent the location of these changes in *A. azarai*, whereas colored bars represent transcriptional changes in the other taxa examined, with the key for these taxa shown in the key above the image. The interspecific clustering of these changes would indicate the localization of amino acid variation at putatively functional portions of the mature *AVPR1A* mRNA. The *x*-axis represents the location of those changes along the first exon of *AVPR1A*, with the ATG situated at the graph's origin (0). **Panel b.** A two-dimensional projection of non-synonymous amino acid changes specific to *A. azarai AVPR1A* coding sequence, as derived from the codon position variance histogram (**Figure 4.7a**) and simulated in Topo2 (adapted from Fink et al. 2007).

In primates, we see a wide range of variation in the *AVPR1A* coding region (**Figure 4.8**). However, the repetitive GACA and GTATATAC sequences situated in the 5' region are invariant in our *A. azarai* population whereas many species exhibit allelic variation in the 5' upstream region. This finding could imply that purifying selection has acted on the regulatory and coding regions of this gene to conserve the expression and structure of V1a receptor proteins in owl monkeys. Alternatively, the unique evolutionary history of the owl monkey genome may have shaped the pattern of molecular diversity that we observe.

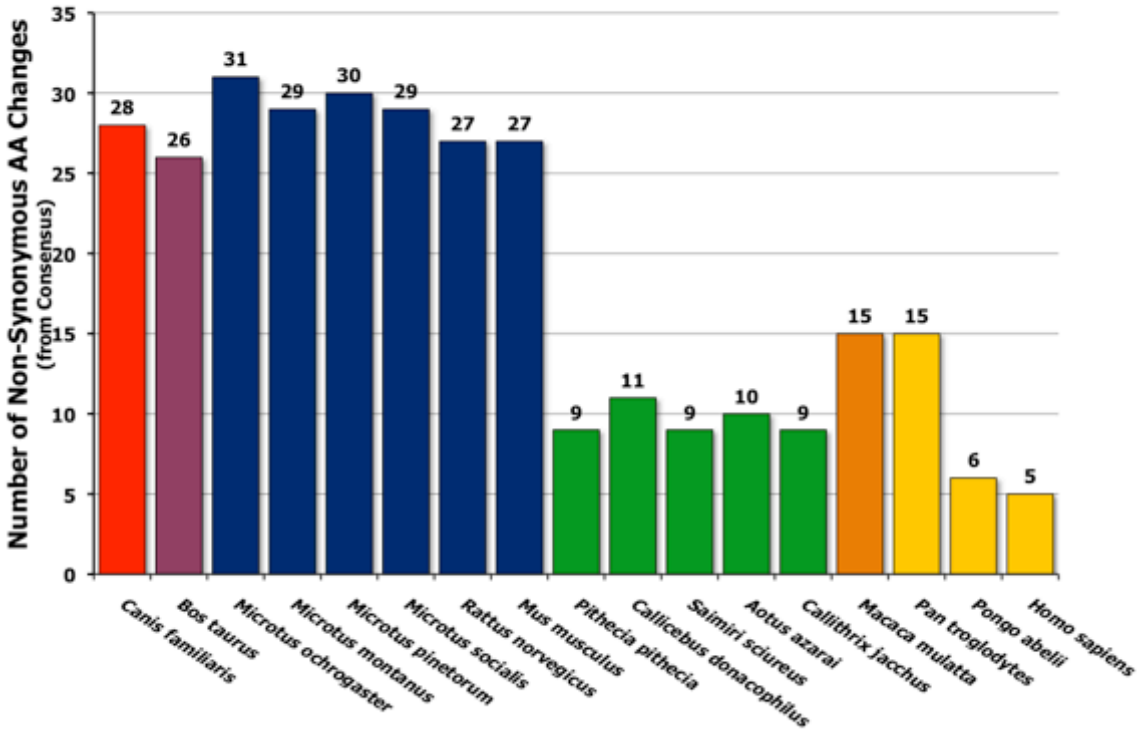


Figure 4.8. The frequency of non-synonymous amino acid changes in the *AVPR1A* gene, as calculated from the consensus translation (by plurality), in 17 mammalian and primate sequences.

Our ability to detect the binding of transcription factors to this homogeneous GTATATAC repeat *in silico* also points to another level of functional evolution. This analysis identified an entire suite of palindromic repeats which would potentially allow for over 100 TATA-like transcription factors to uninterruptedly bind upstream of the *AVPR1A* TSS (**Figure 4.3**). If this motif is, indeed, related to owl monkey *AVPR1A* expression, then it could mean that, given the presence of these general-tissue TATA-like factors, any of the three classes of RNA polymerase could be recruited to the area and initiate transcription (Bell et al., 2001). In this manner, the presence of such an element could act to enhance or insulate *AVPR1A* transcriptional activity in owl monkeys. Regardless of its functional outcome, the presence of this GTATATAC repeat upstream of the RS1 implies that different promoter motifs exist among primates. Furthermore, if its sequence does influence the evolution of different social behaviors, then it is not simply an issue of whether these kinds of repeats are present or absent in a given species.

Based on these findings, it is likely that the function of the *AVPR1A* locus in *A. azarai* is influenced by variation that occurs at the coding, expression, and structural levels. This is a very different scenario than would be expected for the evolution of other GPCR-class proteins which, in the recent comparison of six mammalian genomes, were denoted as a class of conserved, positively selected genes (Kosiol et al., 2008). Instead, different genic components could independently experience distinct degrees of sequence conservation or plasticity in different evolutionary lineages, depending on finely tuned selective processes acting upon those organisms in different ecological contexts. In fact, recent work on *AVPR1A* coding variation in deer mice (*Peromyscus* spp.) suggests that different portions of the amino acid sequences are less conserved than others (Turner et

al., 2010). This observation was also made in studies of voles with regard to the N-terminus of the mature V1a receptor protein (Fink et al., 2007), and is reflected in our own examination of owl monkey *AVPR1A* coding region sequence diversity.

AVPR1A evolution in platyrrhines

Platyrrhine primate species have accumulated many synonymous and non-synonymous mutations in the coding sequence of *AVPR1A* when compared to Old World monkeys and hominoid primates. Each of the six platyrrhine genera examined in this study (*Aotus*, *Lagothrix, Saguinus*, *Callithrix*, *Pithecia* and *Saimiri*) exhibits a number of lineage-specific mutations that distinguishes their *AVPR1A* coding sequences from those of any other taxon being investigated. In addition, all platyrrhines share 24 nucleotide changes (of the 563 nucleotides in our interspecies comparison) that separate their clade from that of other primates (**Figures 4.4b**, **4.5a**). Although the *AVPR1A* coding diversity observed in platyrrhines is undoubtedly a result of the colonization of the New World by proto-platyrrhines some 30 million years ago (Poux et al., 2006; Hodgson et al., 2009), it may also reflect the consequences of selection acting upon this locus, especially given the enrichment of monogamous social systems among New World species (Di Fiore and Rendell, 1994; Schwindt et al., 2004; Rendell and Di Fiore, 2007).

That selection may be at work at this locus across the Order Primates is indicated by the statistical analysis of *AVPR1A* coding sequences. The high positive value of Tajima's *D* and negative value of Fu's F_s values reveals an excess of intermediate frequency alleles, which, in turn, suggests that the observed variation has been shaped by balancing selection. Thus, despite its intraspecific coding conservation, as shown in 25

A. azarai individuals, the *AVPR1A* gene exhibits considerable nucleotide and codon diversity among closely related primate species, and the multiple clusters of STRs in its 5' regulatory region present another tier of variation altogether.

While searching for more information on the transcriptional control of expression of *AVPR1A* in primates, we compared our aligned sequences with the genomes of other primate species on UCSC GB (Kent et al., 2002; Karolchik et al., 2008). Our genomic comparison of structural variation at the chromosomal level revealed the conservation of the *AVPR1A* gene in all species being compared, as well as the fact that, in many species, chromosomal rearrangements have apparently reshuffled the regulatory regions upstream of the gene (hg19, chr12: 61826483-61832857) (**Figure 4.9**). In addition, when surveying this region using the DGV (Iafrate et al., 2004) track on UCSC GB, we noted that, for many primate taxa, CNVs exist in large sections of the entire region. Recent genomic studies have noted that simple variation (STRs), similar to those present adjacent to *AVPR1A*, may precipitate more mutations, both substitutional and structural. Repetitive sequences that are enriched in promoter regions are also enriched in CNV breakpoints, suggesting that the same properties that enable regulation of transcription may also be mildly mutagenic for the formation of CNVs, and as a consequence, CNVs may influence the evolution of gene regulation (Conrad et al., 2010). Thus, genomic structural variation and STR motifs may both contribute to the apparently increased potential of transcriptional plasticity and differential gene dosages of the *AVPR1A* locus in primates, rodents, and other mammals.

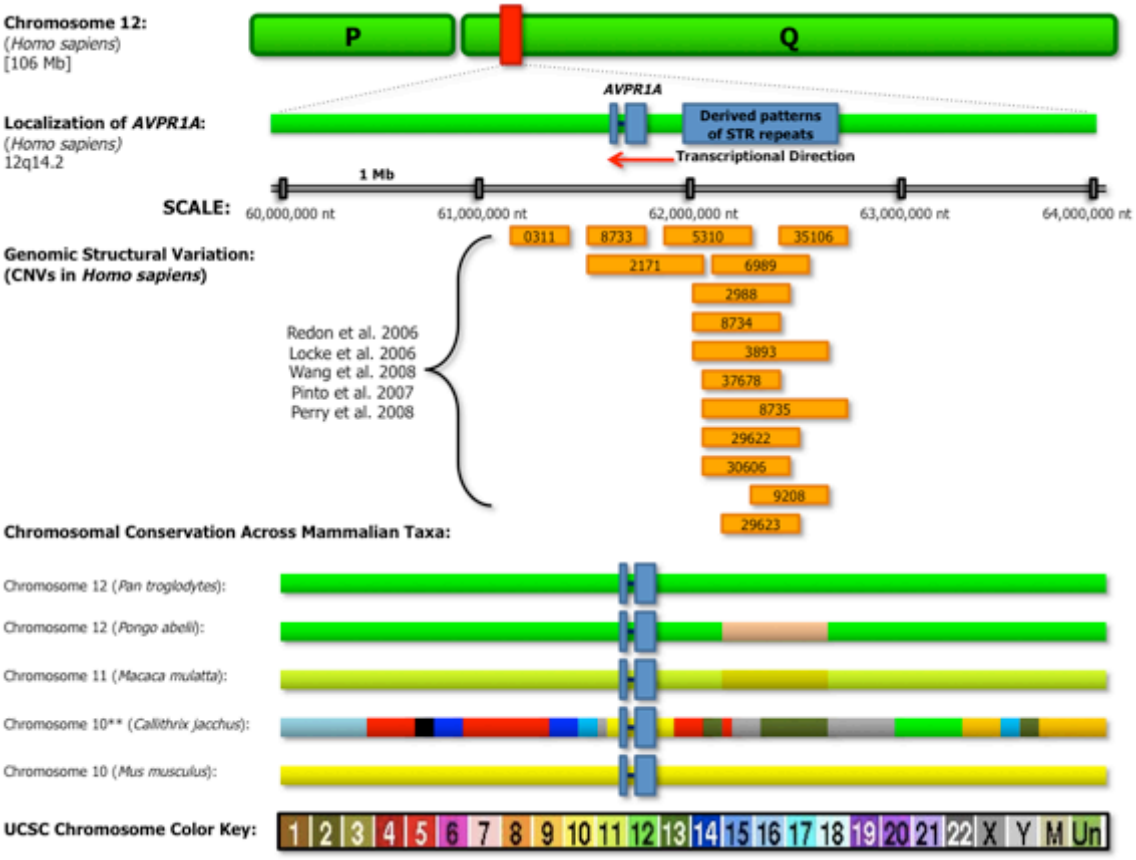


Figure 4.9. Genomic structural variation near the *AVPR1A* gene in different mammalian species. The particular chromosomal location of the gene has changed multiple times throughout mammalian evolution, as shown by the chromosome maps on the lower portion of the figure (adapted from UCSC GB [Kent et al., 2002; Karolchik et al., 2008]).

Along the same lines, the *AVPR1A* 5' regulatory region has been shown to be highly variable, even in closely related taxa (Kim et al., 2002; Hammock and Young, 2005; Donaldson et al., 2008; Rosso et al., 2008). Our data for upstream microsatellite variation in different primate taxa clearly support these findings. This putative promoter region has been empirically linked to the functional expression of the *AVPR1A* gene, as well as the density and distribution of V1a receptors in the brain, yet is tremendously variable across species, and susceptible to major rearrangements and insertion-deletion events. Therefore, it is possible that in primates (and other mammals), the STR-bearing promoter could be under relaxed selection to permit generational expression changes to

cope with changing social or ecological conditions. Alternatively, the molecular structure of the region could simply be conducive to accelerated mutability, taking the form of SNP and in-del substitutions, STR length variants, non-allelic homologous recombination, and even CNV formation.

Role of AVPR1A in the evolution of primate sociality

Ecological conditions may be a strong selective influence on the evolution of primate social behavior, and molecular sequence changes could serve as a means of adapting to these ecological challenges. Regulatory regions provide additional locations for mutations to accumulate, and, in some cases, could facilitate quick adaptive responses (via gene transcriptional regulation) to environmental change. Therefore, a certain level of plasticity in these regions might be allowed to exist within populations. Moreover, primates may, indeed, possess noteworthy levels of inter- and intraspecific variation in their *AVPR1A* coding regions (Conrad et al., 2010).

Many molecular mechanisms can reshape genetic sequences and potentially alter gene dosage and function. CNVs and STRs are present throughout the *AVPR1A* region in mammals (Kent et al., 2002; Karolchik et al., 2008), and could facilitate changes in gene expression in different species or even individuals (Gökçümen and Lee, 2009; Conrad et al., 2010; Park et al., 2010). Therefore, the careful investigation of both coding and non-coding regions for signatures of selection may be essential for elucidating the evolution of partner preference and pair bonding in extant primate species.

In conclusion, we have characterized the molecular features of *AVPR1A* in *A. azarai* and other platyrrhines, and demonstrated that its sequence differs from that seen at

this locus other primate species. These data provide new clues about the possible basis of pair bonding in New World species, and may help to explain the sporadic appearance of monogamy in this infraorder. This study also reinforces the notion that primates have experienced significant neurogenetic variation during their evolution, and suggests that monogamy has arisen multiple times in the primate order through different molecular mechanisms.

Chapter 5: *OXTR* Variation

Summary[9]

The oxytocin (OT) hormone pathway is involved in a multitude of physiological processes, and one of its receptor genes (*OXTR*) has been implicated in increased partner preference and pair bonding behavior in mammalian lineages. This observation is of considerable importance for understanding social monogamy in primates, which is present in only a small subset of primate taxa, including Azara's owl monkey (*Aotus azarai*). To examine the potential relationship between social monogamy and variation at the oxytocin receptor gene, we sequenced the regulatory (>4900 bp) and coding regions (2 exons transcribing 389 amino acids) of this locus in 125 owl monkeys from the Guaycolec Estancia study population. We also assessed the interspecific variation of *OXTR* by sequencing the locus in 25 additional primate and rodent species that represent a set of phylogenetically and behaviorally disparate taxa.

The resulting data revealed *A. azarai* to have a unique genic structure for *OXTR* that varies in its coding sequence relative to other primate and mammalian species. Furthermore, the *A. azarai* sequence exhibits four amino acid changes that may putatively increase the surface area of the ligand-binding domain. We also found that, in general, platyrrhine primates possess more lineage-specific nucleotide substitutions and non-synonymous amino acid changes in their *OXTR* coding sequences than do their catarrhine and hominoid relatives. Such coding diversity could be associated with the novel form of OT (leucine→proline at amino acid residue eight [8]) that is found in New World monkeys (Lee et al., 2011).

[9] **The results presented in this chapter will form the basis of a new academic report:**
Babb PL, Fernandez-Duque E, Schurr TG. (*In Preparation*). Monogamous owl monkeys differ from other non-monogamous primates in the structure of *OXTR*.

Yet, the forces of purifying (negative) selection appear to be at work on *OXTR* protein coding sequences in primate lineages. For all primate taxa, silent nucleotide substitutions (leading to synonymous amino acids) vastly outnumber mutations that lead to non-synonymous amino acid changes. By contrast, the split between rodents and primates is characterized by many non-synonymous changes (including one codon deletion), suggesting that diversifying (positive) selection may have played a role in the creation of different primate *OXTR* proteins.

Regulatory regions of *OXTR* also vary among primates, with large STR expansions being inserted in the *OXTR* 5' UTR/promoters of *Pan troglodytes* and *Homo sapiens*. This pattern is in marked contrast with that for the 5' UTRs of platyrrhines such as *Aotus* and *Callithrix*, whose promoters are peppered with more than 10 deletions of 5 nucleotides (or larger), including a 30 bp deletion that contains the translational start site of *OXTR* in all other primate genomes analyzed. These Homininae-specific insertions and Platyrhinni-specific deletions could greatly affect the transcriptional efficiency of the *OXTR* gene in those taxa. Alterations of expression would likely reduce or increase *OXTR* mRNA translation, and subsequently influence the distribution and density of *OXTR* proteins in the brain. It is conceivable that such changes could, in turn, have dramatic effects on signaling capabilities of the OT pathway, and ultimately impact phenotypic behavior.

5.1 Introduction

Monogamous social systems are rare among primates, yet are widely distributed across the order, indicating that they have arisen independently multiple times throughout primate evolution (van Schaik and van Hoof, 1983; van Schaik and Dunbar, 1990). Although "monogamy" necessarily involves many different complex social behaviors (Mendoza et al., 2002; Moller, 2003), this scenario raises an important question: are the social behaviors of primates controlled by the same genes in different species? By comparing genetic regions that are related to these behaviors in different primate taxa, one can begin to explore whether the evolution of complex social behavior occurs through conserved molecular mechanisms.

There is growing evidence that certain neurogenetic loci play key roles in the expression of social and affiliative behaviors in mammals. For example, the neuropeptide hormone oxytocin (OT) and its neuron-based receptor, *OXTR,* have been linked to mammalian social behaviors and emotional states that promote sociality (Carter, 1998; Carter et al., 2008). It is therefore possible that lineage-specific mutations in *OXTR* may have induced changes to the OT pathway, and influenced the differential evolution of sociality in mammals, including primates.

Monogamous social systems are particularly enriched within New World monkeys (Kleiman, 1977; Rendall & Di Fiore, 2007). Azara's owl monkeys (*Aotus azarai*) from Formosa, Argentina, exhibit many of the behaviors associated with monogamous social systems (Fernandez-Duque, 2009). These behaviors include temporally pervasive pair bonds between adults, biparental care of infants, territoriality, and even food sharing (Wright, 1994; Rotundo et al., 2005; Wolovich et al., 2008).

Given these behavioral attributes, this species represents an ideal model in which to investigate the role of neurogenetic loci in influencing social and mating behavior in primates.

Like arginine vasopressin (see Chapter 4: *AVPR1A*), OT is pleiotropic, multifunctional, and synthesized in the hypothalamus. However, OT transmits signals to its receptors via osmotic cellular diffusion across cell membranes, whereas AVP conveys its signals through synaptic connections between neurons (Carter et al., 2008). Nevertheless, current evidence suggests a common molecular origin for the AVP and OT pathways in mammals.

The amino acid sequences of the AVP and OT hormones, as well as their receptors, *AVPR1A* and *OXTR*, show a high level of similarity, and coalesce to common ancestral proteins in phylogenetic reconstructions (Carter et al., 2008). It is believed that both pathways derive from the hybrid/ancestral vasotocin pathway present in fish and reptiles (Gwee et al., 2008), and that the receptor genes *AVPR1A* and *OXTR* represent the duplication and diversification of gene function during mammalian evolution (Carter et al., 2008).

In its role as a signal transmitter, OT coordinates social behaviors within the neuroendocrine network, where it responds to the activity of various stressors, usually by reducing reactivity or promoting withdrawal (Grippo et al., 2008, 2009). AVP also works within this network, but more frequently generates or influences arousal, vigilance and defensive behaviors (Carter, 1998; Carter et al., 2008). Thus, the two hormones have exerted their effects through different sorts of rewarding or protective behaviors (Insel, 2003; Edwards & Self, 2006), and operate through independent cellular mechanisms.

OT plays an important role in positive social interactions, such as grooming, sex, childbirth and parental care (Carter et al., 2008), and has also been associated with social signaling (Bales et al., 2007a, b, c). For example, touch, orgasm, the provision of social support, and even the presence of conspecifics have been shown to stimulate the release of OT in different species. In addition to its apparent responsiveness to social signals, OT also seems to plays a role in the orientation toward, and processing of different social signals. In fact, deficits in the processing of social information have been linked to low OT levels (Takayanagi et al., 2005). In particular, it has been suggested that the social deficits in autism might be related to low levels of AVP and OT (Hammock & Young, 2006). Hence, the neuroendocrine OT pathway is integral to developing and maintaining social relationships between individuals.

As such, both OT and AVP have been implicated in long-term affiliative behaviors that are associated with social interactions. These behaviors include huddling, aggression reduction, mating and the formation of temporally pervasive social bonds. For example, studies of voles and other rodents have shown that the pair bonding and offspring investment behaviors that typify affiliative social interactions between individuals are not manifested when the actions of OT and AVP are blocked (Takayanagi et al., 2005; Egashira et al., 2007). Although research has implicated OT and AVP in pair bonding behaviors in primates (Babb et al., 2010; Donaldson et al., 2008; Rosso et al., 2008; Walum et al., 2008), the degree to which variation in these molecular pathways correlates with platyrrhine social systems remains unclear.

Here, we characterize *OXTR* sequence diversity in an owl monkey population, and reconstruct the molecular history of the *OXTR* locus in *Aotus* and other primates. To

assess intraspecifc variation at the *OXTR* locus, we sequenced the entire coding and regulatory regions of *OXTR* in our study population of owl monkeys. To investigate interspecific variation of the *OXTR* mRNA, we compared 26 species of primates and rodents to examine amino acid changes that distinguish platyrrhines and *Aotus* from other primate taxa. Similarly, we explored the structure of the *OXTR* regulatory region of eight haplorhine primate genomes to highlight potential interspecific transcriptional variants.

5.2 Methods

Study population

Azara's owl monkeys, inhabit the gallery forests and the patchy forest-islands interspersed throughout the savannahs of northeastern Argentina. Their social groups are small (two to six individuals), being comprised of an adult heterosexual pair, one infant, and one or two juveniles, and occasionally a subadult (Fernandez-Duque et al., 2001). The adult male in the group is heavily involved in the care of the young, and carries the infant most of the time (84%) after the first week of life (Rotundo et al., 2005). This pattern of social monogamy and paternal care distinguishes owl monkey males from those of the majority of other primate species.

Samples

We assembled a panel of 25 *A. azarai* individuals for the analysis of intraspecific variation of *OXTR* regulatory and coding sequences (**Table 5.1**). The *A. azarai* individuals in the panel represent the most disparate mtDNA (maternal) haplotypes in our study population living in Formosa Province, Argentina (Babb et al., 2010, 2011). Upon

capture, each of the 20 wild individuals underwent a thorough physical exam (Fernandez-Duque and Rotundo, 2003) during which we obtained biological samples for use in genetic analyses.

For comparative analyses of genetic variation at the *OXTR* locus, DNA samples from three putative *Aotus* species and subspecies (*A. nancymaae*, *A. lemurinus*, and *A. nigriceps*) were provided by the Zoological Society of San Diego. Another five *A. nancymaae* samples were obtained from individuals at the DuMond Conservancy for Primates and Tropical Forests (Miami, FL). In addition, we analyzed samples from several other platyrrhine taxa, including two individuals each from titi (*Callicebus donacophilus*), saki (*Pithecia pithecia*), and squirrel (*Saimiri sciureus*) monkeys[10], which we also obtained from the Zoological Society of San Diego. Two human DNAs were also included as positive controls (Hs001: P. Babb, Hs002: Promega pooled DNA). Furthermore, we purchased comparative samples for 13 other primate taxa from the NIA and IPBIR cell line collections, curated by the Coriell Institute of Biomedical Research (Camden, NJ) (**Table 5.1**).

[10] Like owl monkeys, titi and saki monkeys exhibit socially monogamous behavior, whereas squirrel monkeys are known to live in multimale-multifemale groups that vary in their tenure and complexity (Di Fiore and Rendall, 1994; Boinski, 1999; Rendall and Di Fiore, 2007).

Table 5.1. Samples investigated at the *OXTR* locus.

ID	Species	Common Name	Sex	Locale
AaaPLunka	*Aotus azarai*	Azara's Owl Monkey	Unk	Core area, Formosa, AR
Aaa008	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa014	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa015	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa021	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa032	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa034	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa037	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa053	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa057	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa063	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa067	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa071	*Aotus azarai*	Azara's Owl Monkey	M	Downstream, Formosa, AR
Aaa082	*Aotus azarai*	Azara's Owl Monkey	M	Downstream, Formosa, AR
Aaa087	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa092	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa108	*Aotus azarai*	Azara's Owl Monkey	F	Upstream, Formosa, AR
Aaa109	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa114	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa122	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa123	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
AaaF1	*Aotus azarai*	Azara's Owl Monkey	F	Saenz-Peña Zoo, AR
AaaF1B	*Aotus azarai*	Azara's Owl Monkey	F	Saenz-Peña Zoo, AR
AaaF2	*Aotus azarai*	Azara's Owl Monkey	F	Saenz-Peña Zoo, AR
AaaM2	*Aotus azarai*	Azara's Owl Monkey	M	Saenz-Peña Zoo, AR
Al001	*Aotus lemurinus*	Grey-bellied Owl Monkey	Unk	San Diego Zoo/CRES
Ana001	*Aotus nancymaae*	Nancy Ma's Owl Monkey	Unk	San Diego Zoo/CRES
Ani001	*Aotus nigriceps*	Black-headed Owl Monkey	Unk	San Diego Zoo/CRES
Cd001	*Callicebus donacophilus*	White-eared Titi Monkey	M	San Diego Zoo/CRES
Cd002	*Callicebus donacophilus*	White-eared Titi Monkey	M	San Diego Zoo/CRES
Ppp001	*Pithecia pithecia*	White-faced Saki Monkey	M	San Diego Zoo/CRES
Ppp002	*Pithecia pithecia*	White-faced Saki Monkey	F	San Diego Zoo/CRES
Sss001	*Saimiri sciureus*	Common Squirrel Monkey	M	San Diego Zoo/CRES
Sss002	*Saimiri sciureus*	Common Squirrel Monkey	F	San Diego Zoo/CRES
Ana002	*Aotus nancymaae*	Nancy Ma's Owl Monkey	F	DuMond Conservancy
Ana003	*Aotus nancymaae*	Nancy Ma's Owl Monkey	F	DuMond Conservancy
Ana004	*Aotus nancymaae*	Nancy Ma's Owl Monkey	F	DuMond Conservancy
Ana005	*Aotus nancymaae*	Nancy Ma's Owl Monkey	F	DuMond Conservancy
Ana006	*Aotus nancymaae*	Nancy Ma's Owl Monkey	F	DuMond Conservancy
Hs001	*Homo sapiens*	Human	M	Self (P.Babb)

ID	Species	Common Name	Sex	Locale
Hle001	*Hylobates leucogenys*	White-cheeked Gibbon	M	Coriell Institute (IPBIR: PR01038)
Hsy001	*Hylobates syndactylus*	Siamang	F	Coriell Institute (IPBIR: PR00969)
Ppa001	*Pan paniscus*	Bonobo (Pygmy Chimpanzee)	F	Coriell Institute (IPBIR: PR00092)
Ptr002	*Pan troglodytes*	Chimpanzee	F	Coriell Institute (IPBIR: PR00605)
Pab001	*Pongo pygmaeus abelii*	Sumatran Orangutan	M	Coriell Institute (IPBIR: PR00253)
Cge001	*Callitrhix geoffroyi*	White-fronted Marmoset	F	Coriell Institute (IPBIR: PR01094)
Papio001	*Papio anubis*	Olive Baboon	M	Coriell Institute (IPBIR: PR00036)
Cmo001	*Callicebus moloch*	Dusky Titi Monkey	M	Coriell Institute (NIA: NG06115)
Epa001	*Erythrocebus patas*	Patas Monkey	F	Coriell Institute (NIA: NG06116)
Lla001	*Lagothrix lagotricha*	Woolly Monkey	F	Coriell Institute (NIA: NG05356)
Mamu001	*Macaca mulatta*	Rhesus Macaque	F	Coriell Institute (NIA: NG06249)
Sfu001	*Saguinus fuscicollis*	White-lipped Tamarin	F	Coriell Institute (NIA: NG05313)
Sss003	*Saimiri sciureus*	Squirrel Monkey	F	Coriell Institute (NIA: NG05311)
Hs002	*Homo sapiens*	Human	Unk	Promega Control DNA (pooled)

Note: Core study area is located in Formosa Province, Argentina (Lat. = 25º, 59.4 min South; Long. = 58º, 11.0 min West).
[a] = Control *A. azarai* sample used in this study.

We also conducted searches of the publically available genomes accessible through the UCSC Genome Browser and the ENSEMBL browser for the extraction of annotated primate and rodent *OXTR* sequences (Kent et al., 2002; Karolchik et al., 2008). These searches yielded the entire *OXTR* locus for 10 primate and rodent species (**Table 5.2**). In total, we analyzed 25 *A. azarai* individuals for our intraspecific analyses, and 40 individuals (inclusive of AaaPLunk) representing 26 different primate and rodent species (and subspecies) for our interspecific analyses.

Sequencing

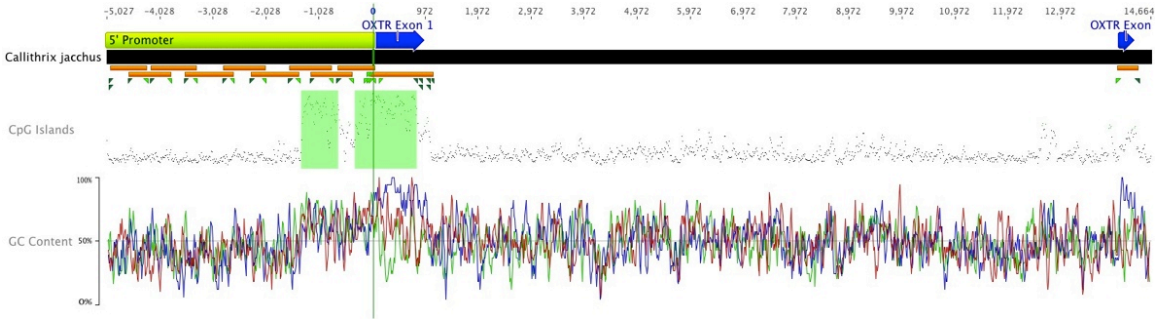
To investigate population-level intraspecific variation at the *OXTR* locus, we targeted the 1.5 kb coding region of the *OXTR* gene, along with >5 kb of adjacent non-coding sequence (Inoue et al., 1994; Kent et al., 2002). These regions included: 3.4 kb of

Table 5.2. Comparative samples used in *OXTR* sequence analyses.

ID	Species	Common Name	Chromosomal Location of *OXTR*	Genome Browser/Portal
hg19	*Homo sapiens*	Human	3	UCSC Genome Browser
panTro3	*Pan troglodytes*	Chimpanzee	3	UCSC Genome Browser
gorGor3	*Gorilla gorilla gorilla*	Gorilla	3	UCSC Genome Browser
nomLeu1	*Nomascus leucogenys*	Gibbon	unknown	UCSC Genome Browser
ponAbe2	*Pongo pygmaeus abelii*	Orangutan	3	UCSC Genome Browser
rheMac2	*Macaca mulatta*	Rhesus Macaque	2	UCSC Genome Browser
calJac3	*Callithrix jacchus*	Marmoset	15	UCSC Genome Browser
mm9	*Mus musculus*	Mouse	6	UCSC Genome Browser
rn4	*Rattus norvegicus*	Rat	4	UCSC Genome Browser
otoGar3	*Otolemur garnettii*	Bushbaby	unknown	Ensembl

5’ flanking sequence upstream of the transcriptional start site (TSS), the entire 1.5kb of the 5’ untranslated region (UTR), the two exons that encode the *OXTR* protein (1170 bp), a small sampling of 36 bp situated at the 3’ end of the 13 kb intronic region separating the exons, and 130 bp from the 3’ UTR directly following the stop codon (TGA) at the end of exon 2 (**Figure 5.1**).

Panel a.



Panel b.

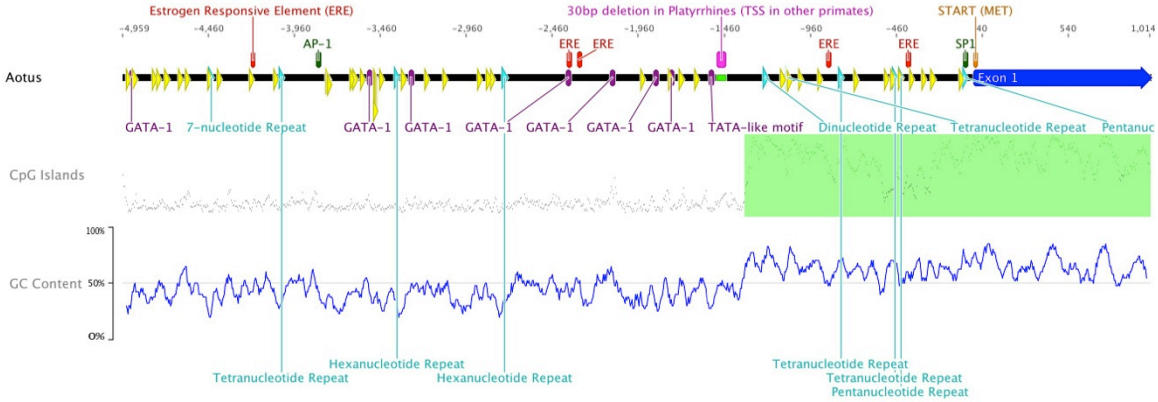


Figure 5.1. Panel a. A schematic diagram of the *OXTR* locus (20 kb window) in the marmoset genome (calJac3), shown 5'-3' left to right. Notable architectural landmarks of this gene are depicted on the figure. The locations of the oligonucleotide primers used in this study for PCR amplification are shown in their approximate positions (depicted as light (F) and dark (R) *green* triangles) given the diagrammatic scale, as well as the coverage achieved by using different primer pairings for DNA amplification and sequencing (*orange* bars). The 5' UTR (a.k.a. promoter) is shown in *yellow-green*, and the two exons coding for the *OXTR* mRNA are shown in *blue*. CpG islands and GC (%) content are represented as parallel tracks beneath the main figure. **Panel b.** The structure of the *OXTR* regulatory region (5' UTR, promoter) in *A. azarai*. The localization of putative transcription factor binding sites are denoted by *yellow* triangles. Estrogen response elements are annotated in *red* as "ERE", repetitive regions are marked in *blue*, and GATA-1 and TATA-motifs are shown in *purple*. CpG islands and GC (%) content are represented as parallel tracks beneath the main figure.

In order to directly target these regions, we generated a >20 kb interspecific *OXTR* genic contig from genomic sequence files retrieved from UCSC GB, and ENSEMBL using Sequencher, v4.9 (Gene Codes) and Geneious Pro v5.5.5 (Drummond et al., 2010). From this assembly, we designed 28 unique *OXTR* primers to bind in regions exhibiting high levels of sequence conservation across all taxa, using the oligo

software programs NetPrimer (Premier BioSoft) and Primer3 (Rozen and Skaletsky, 2000). These primers amplified 10 overlapping fragments of ~700 bp in length, with alternate primers situated within amplicons to facilitate sequencing (**Table 5.3, Figure 5.1**).

Table 5.3. Primer sequences designed and optimized for *OXTR* amplification and sequencing.

Genetic Region	Primer Name	Strand	Oligonucleotide Sequence (5'-3')
5' Upstream Region (Long-range fragment PCR)	LR01-F	Forward	CTC AGC AAA CTA ACA CAG GAA CAG
	LR01-R	Reverse	GGC GAT GCT CAG GTG CTT C
	LR02-F	Forward	GGA GAG CAT CAG GAA AAA TAG C
	LR02-R	Reverse	GCC CAT AGA AGC GGA AGG TG
	LR03-F	Forward	TGA CCT CCC TCA GCA AGA AG
	LR03-R	Reverse	CAG GTA GGT GGA GGC GAA C
5' Upstream Region (Overlapping coverage of 5' UTR)	Reg01-F	Forward	GAA CGC ACC CTC CAT GAC CC
	Reg01-R	Reverse	CAC CGA AGC AGG TAC TGT GGA
	Reg02-F	Forward	TTC CTC CGG GAC TGG GAA TC
	Reg02-R	Reverse	ATT CCA ACC GAG GCT CCA GA
	Reg03-F	Forward	AAC GTT CGG GAA ACC TCG AC
	Reg03-R	Reverse	CTG GAG GTG TGG GAG GAG AG
	Reg04-F	Forward	CTG CAA GAG GGA AGG AAC TCG T
	Reg04-R	Reverse	GGA CTT GCC CAA GGT CAC TCC A
	Reg05-F	Forward	AGC CAA GGA CCA GGT TTC CT
	Reg05-R	Reverse	TTC CTG CCT GGT TTC TGC AC
	Reg06-F	Forward	AGC CTG TGA GAC AGC CTA ACT
	Reg06-R	Reverse	GGT ATC CAT GGG GGA CTG GT
	Reg07-F	Forward	TGC CAT GTG CCA ACT ACA AC
	Reg07-R	Reverse	ACT GAC CTC CCT CAG CAA GA
	Reg08-Fb	Forward	AGA TGG GGT GCC TGG AAG GAC T
	Reg08-Rc	Reverse	GGG AAC AAC ACA CAG TGG GG
	Reg09-F	Forward	TCT CAT GCA AAA ACA ACC TCA CAG A
	Reg09-Ra	Reverse	TGT CTT TTG ACC CAG CAA TCT C
Coding Region (Exon 1)	Ex1-F2b	Forward	CCT ACA CCC TCC GAC AC
	Ex1-R3b	Reverse	GAC TCT GTG GGA TTT CAA AC
Coding Region (Intron, Exon 2, 3' UTR)	Ex2-F	Forward	TGC CCT TTT CTC CTT TCT CC
	Ex2-R	Reverse	CCT ACT GTA GCC ACC CCA AG

PCR cycling parameters were optimized for each primer pair using a Touchgene Gradient thermocycler (Techne), and all subsequent reactions were PCR amplified in GeneAmp 9700 thermocyclers (ABI). Amplified products were visualized on 1% TBE SeaKem agarose (Lonza) via gel electrophoresis.

Amplicons were purified and DNA sequences were read on a 3130xl Gene Analyzer, as described in previous chapters. We assessed read quality for each sequence using Sequencing Analysis v5.4 (ABI) and aligned them using Sequencher and Geneious. This approach generated >6.2 kb of reliable *OXTR* genic sequence for each individual.

Sequence alignments

For each individual, we assembled sequences for the distinct genic regions of *OXTR* by aligning the overlapping forward and reverse fragments using strict (95%) agreement thresholds. To investigate variation present in owl monkeys and other primates across the *OXTR* locus, we constructed separate sequence matrices to examine the nucleotide diversity in five regions of differential genic functionality for all individuals. These matrices isolated sequence variation within 5' UTR (5899 bp in putative promoter), Exon 1 (922 bp), Intron (39 bp), Exon 2 (248 bp), and 3' UTR (142 bp) regions in both our intra- and interspecific analyses. In addition, we obtained and assembled coding sequences representing the entire mRNA of the *OXTR* gene (Exons 1 + 2=1170 bp: 389 amino acids + TGA stop codon) in all *A. azarai* individuals, and compared the structure of the owl monkey receptor protein to all comparative samples.

Using this collection of sequences, we searched for any polymorphic mutations and putative transcription factor binding sites (TFBS) in the upstream promoter of the *OXTR* gene in *A. azarai* relative to other primate species. This study was conducted using the MatInspector software package (GenoMatix Software, GmbH), with parameters being set for transcription factors present in mammalian neuronal cellular tissues as derived from human and mouse consensus databases. This work was later confirmed

using the transcription factor prediction software plugin for Geneious v5.5.5 (Drummond et al., 2010).

Population genetic analyses

Summary statistics, including pairwise identity, mean number of pairwise differences (π), GC content, and haplotype (h) diversity, were calculated for all *OXTR* sequence matrices using programs available in Arlequin v3.5 (Excoffier et al., 2005), Geneious v5.5.5 (Drummond et al., 2010) and DnaSP v4.50 (Rozas et al., 2003). We also calculated the neutrality indexes of Tajima's *D* (Tajima, 1989a, b) and Fu's F_s (Fu, 1997) to explore the demographic and evolutionary forces possibly acting on the coding region of *OXTR* in our study population. Transversions (TV) were weighted higher than transitions (TI) (10:1 final ratio) in all statistical calculations to account for the molecular probability for either occurring across evolutionary time (Bandelt et al., 1995, 1999, 2000, 2002).

Phylogenetic analyses: Data alignment

For phylogenetic analyses of *OXTR* sequences from platyrrhine and other primate species, we restricted the number of *A. azarai* sequences used in the analyses to one individual (AaaPLunk), which possessed the most frequent haplotype (identified in our study population, Babb et al., 2011) for each of the five delineated genic regions (5'UTR, Exon 1, Intron, Exon 2, 3'UTR). This decision was made to avoid biasing the range of sequence variation toward the species *A. azarai*.

For each of the five predefined genic regions, we aligned the AaaPLunk *A. azarai*

sequence with the appropriate sequences derived from our 29 comparative samples, as well as with the 10 primate and rodent *OXTR* sequences obtained from the UCSC GB and ENSEMBL databases. The resulting five interspecific matrices, each consisting of 40 sequences representing 26 species and subspecies, were pruned to avoid the ambiguous characters (*N*) found in some of the public genomes, and were carefully edited to account for insertion-deletion polymorphisms. The comparative dimensions of each matrix, and the number of useable sites for each genetic region can be found in **Tables 5.4** and **5.5**.

Table 5.4. Summary statistics for *OXTR* sequences in 25 Azara's owl monkeys (*A. azarai)*.

	5' UTR 4936 bp	**Exon 1 922 bp**	**Intron 36 bp**	**Exon 2 248 bp**	**3' UTR 130 bp**
Sequence diversity:					
Pairwise identity of alignment (%)	99.96%	99.95%	100%	99.80%	99.97%
GC content (% non-gap)	47.60%	64.80%	63.90%	59.50%	60.00%
Summary statistics:					
Usable loci (<5% missing data)	4936	922	36	248	130
Identical sites	4930 (99.9%)	921 (99.9%)	36 (100%)	246 (99.2%)	129 (99.2%)
Polymorphic sites	6	1	0	2	1
Single nucleotide polymorphisms (SNPs) [freq.]	314C>G (4%) 1250A>G (2%) 2537G>A (30%) 3213C>T (30%) 3456A>G (24%) 4319G>A (32%)	819T>A (34%)	0	133T>C (8%) 197G>C (26%)	127T>A (2%)

Table 5.5. Summary statistics for *OXTR* sequences in 40 primate and rodent samples (26 species and subspecies).

	5' UTR†† 5896 bp*	Exon 1 922 bp	Intron 39 bp*	Exon 2 248 bp	3' UTR 142 bp*
Sequence diversity:					
Pairwise identity of alignment (%)	87.60%	94.90%	90.90%	94.80%	84.00%
GC content (% non-gap)	41.50%	65.30%	59.10%	59.40%	58.80%
Nucleotide diversity:					
Nucleotide diversity (avg. over loci)	0.29 +/- 0.16	0.21 +/- 0.10	0.33 +/- 0.17	0.17 +/- 0.08	0.54 +/- 0.26
Summary statistics:					
Usable loci (<5% missing data)	5986	922	39	248	142
Identical sites	3882 (65.8%)	665 (72.1%)	22 (56.4%)	183 (73.8%)	28 (19.7%)
Polymorphic sites	2014	257	17	65	114
Sites with transitions (ts)	611	175	9	49	55
Sites with transversions (tv)	369	116	6	24	48
Insertion/deletions (indels)	1156	3**	6	0	62
Selective neutrality:					
Mean # pairwise differences (π)	1350.64	191.47	12.29	42.33	70.23
Sites with substitutions (S)	933	250	13	64	84
Tajima's *D* (1000 simul.)	15.14	8.41	9.42	6.54	9.31
P (*D* simul < *D* obs)	0	0.008§	0.17	0.006§	0.02
Fu's F_S (1000 simul.)	4.1	-3.79	-24.44	-14.44	-8.78
P (sim_F_S ≤ obs_F_S)	0	0	0	0	0

†† = 8 haplorhines were analyzed at full resolution (>5 kb) for calculation of summary statistics of *OXTR* 5' UTR region.
* = Indicates width of nucleotide matrix including alignment gaps (to account for insertion/deletion events).
** = The 3 indels reported here constitute a rodent-specific 3 bp deletion of codon number 253, as observed in mm9 and rn4.
§ = *P* value suggests statistical significance.

The resulting matrices were annotated in Geneious (Drummond et al., 2010), then exported as Nexus (.nex) files for appropriate command-block formatting for phylogenetic tree calculations in PAUP* v4.0b10 (Swofford, 2002), PAML/codeML (Yang, 2007), and MrBayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The output files from these analyses were imported into FigTrees v1.3.1 (A. Rambaut, http://beast.bio.ed.ac.uk/FigTree) for tree visualization and phylogram generation.

Phylogenetic analyses: Network analysis

We generated multistate MJ networks from all *A. azarai OXTR* sequences using Network v4.5.02 (Bandelt et al., 1995, 1999, 2000, 2008). We again employed a 1:10 ts:tv weighting scheme (Bandelt et al., 1995, 1999, 2000) with slight modifications for the construction of networks in order to reduce the phylogenetic reticulations caused by homoplasies and hypervariable characters.

We similarly produced a multistate amino acid MJ network for *OXTR* mRNA sequences from comparative primate and rodent taxa to explore the degree of amino acid variation in the gene across broader evolutionary distances. To evaluate this level of molecular diversity, we translated the *OXTR* amino acid sequences from the 1170 bp mRNA sequence matrix (combined Exon 1 + Exon 2, minus the TGA stop codon) in Geneious, and ran the resulting 40 sequences of 389 amino acids in Network.

Phylogenetic analyses: Model selection

To select the most appropriate model for our phylogenetic analyses of the complete mRNA sequence of *OXTR*, we ran the program jModelTest v0.1.1 (Guindon and Gascuel, 2003; Felsenstein, 2005; Posada, 2008) using 11 substitutions patterns to survey 88 models of nucleotide substitution (+F base frequencies, rate variation of +I and +G with nCat=4). Using our matrix of 40 sequences of 1170 bp, we conducted three parallel searches using the standard Akaike Information Criterion (AIC), BIC, and performance-based DT search parameters. The base tree for our likelihood calculations was optimized for ML phylogenetic analysis.

Phylogenetic analyses: ML and BI analyses

We conducted ML analysis in PAUP* v4.0b10 (Swofford, 2002) to estimate the most likely evolutionary tree based on the alignments of *OXTR* nucleotides and amino acid codons and using the nucleotide model specified by our jModelTest runs. For ML analysis, we estimated bootstrap values based on a set of 10,000 replicates.

To find the phylogenetic tree or set of trees that maximizes the probability of obtaining our data given a specified model of evolution, we undertook Bayesian inference analysis (BI) with the software program MrBayes v3.1.2 (Huelsenbeck and Ronquist, 2001; Huelsenbeck et al., 2001; Ronquist and Huelsenbeck, 2003). In our BI analysis, we used the ML function and employed the BIC-specific nucleotide substitution model as suggested by jModelTest. The Markov Chain Monte Carlo (MCMC) search was run with four chains for 1,000,000 generations, with trees sampled every 1000 generations. Using the average standard deviation in split frequencies among the four chains, we were able to assess the level of convergence (chain deviation <0.05) denoting an acceptable level of post-convergence tree likelihoods that influence the accuracy of our consensus Bayesian tree. The first 5000 trees were discarded as "burn-in" to remove extraneous pre-convergence probability values that can skew the values of a given consensus Bayesian tree (Huselsenbeck et al., 2001; Altekar et al., 2004).

We also employed MP and ME/NJ tree-building methods to identify any incongruence between the different algorithms. We estimated bootstrap values for MP and ME/NJ trees based on a set of 10,000 replicates.

Phylogenetic analyses: Analysis of adaptive evolution

To identify the level of *OXTR* codon variation, we examined the non-synonymous-to-synonymous amino acid substitution ratio (d_N/d_S) in our mRNA matrix (40 sequences by 1170 bp). Applying the ML tree (identified through likelihood analysis in PAUP*) to the codeML program within PAML (Yang, 2007), we calculated the relative rates of amino acid changes along the different phylogenetic branches. This analysis was undertaken to detect any possible signatures of adaptive evolution and directional selection, such as an excess of non-synonymous mutations along a particular taxonomic lineage. We ran codeML twice to produce two phylogenetic models (M0 [uniform tree rate] and M1 [branch-specific rates]), which were then subjected to a likelihood ratio test (LRT) to assess the significance of estimating individual d_N/d_S ratios (variable) versus estimating one ratio (homogeneous) for the entire tree.

Phylogenetic analyses: Bayesian coalescent estimation

To check for the presence of *OXTR* mutation rate variation among the 26 species and subspecies, and to assess the impact that a relaxed molecular rate of evolution would have on age estimations for hypothetical ancestral *OXTR* gene sequences, we analyzed our 1170 bp mRNA matrix in the software program BEAST v1.5.4 (Drummond et al., 2006; Drummond and Rambaut, 2007). This program generates a range of coalescent events based on prior knowledge of calibrated paleontological events (Drummond et al., 2002; Drummond and Rambaut, 2007). We used the companion software program BEAUti v1.5.4 (Drummond et al., 2006; Drummond and Rambaut, 2007) to apply time points of time to most common ancestor (TMRCA) for specific groups of taxa (taxon

groups and priors outlined in Babb et al., 2011 – see Chapter 2, Table 2.5).

By implementing the Yule speciation process parameters in BEAST, we specified the abovementioned time points as log normally distributed priors applied to the appropriate taxon designations. We used the relaxed-clock lognormal molecular clock and ran the software using the nucleotide model specified by our jModelTest runs. BEAST was run for 2,000,000 generations, echoing on-screen every 10,000 and logging every 200.

The Bayesian coalescent results generated in BEAST were analyzed with the companion software in TRACER v1.5 (Rambaut and Drummond, 2007) to view the distribution of coalescent time points for our prior specified monophyletic taxonomic groups. The 10,001 trees retained from the BEAST run were summarized in TreeAnnotator v1.5.4, and displayed in the program FigTrees v1.3.1.

Genomic comparisons

To investigate the transcriptional control of expression of *OXTR* in primates, we compared our aligned sequences and repeat motifs to the primate genomes on UCSC Genome Browser (panTro3, calJac3, gorGor3, ponAbe2, nomLeu1; Kent et al., 2002; Karolchik et al., 2008), using the *Homo sapiens* (hg19) genome as a reference track. In addition, we applied the most recent annotation of these regions using the Database of Genomic Variants (DGV) (Iafrate et al., 2004) analytical track on UCSC GB to investigate the occurrence of structural variants localized near the *OXTR* locus.

5.3 Results

OXTR sequence diversity in A. azarai

In our analysis of 25 owl monkeys, we generally observed low levels of diversity in the *OXTR* locus, ranging from 99.8% - 100% pairwise identity across the five regional alignments (**Table 5.4**). Nevertheless, diversity was indeed present within the owl monkey population. We detected three SNPs in the coding region of *OXTR* (1170 bp) that occurred at notable frequencies (8%, 26%, and 34% of 50 chromosomes) in the 25 individuals that constitute the *A. azarai* diversity panel. Within the larger population (*N*=111 individuals), these SNPs occurred at even higher frequencies: 8%, 49%, and 65%. The three intraspecific coding region SNPs were transitions, and none led to amino acid changes. Nevertheless, the data revealed *A. azarai* to have a unique genic structure for *OXTR* relative to other primate and mammalian species.

We also identified six SNPs in the 5' upstream regulatory region of *OXTR* (4936 bp). Four of these SNPs were found at high frequencies (24%-32% of chromosomes), whereas two of the 5' regulatory SNPs occurred only in one or two of chromosomes analyzed. None of the intraspecific regulatory variants led to putative transcription factor binding site changes. In addition, we observed zero variation in the small (36 bp) alignment of *OXTR* intronic sequences, and only one occurrence of a single SNP in our 130 bp alignment 3' UTR sequences.

As a result, the ten *OXTR* SNPs (7 regulatory, 3 coding) across the five regions defined 13 distinct haplotypes in owl monkeys. This pattern was confirmed by the results of our intraspecific network analysis (**Figure 5.2a, 5.2b**). Interestingly, although the 5' regulatory region contained a greater number of intraspecific SNPs (6 in 4936 bp), the

coding region contained a higher proportion of intraspecific SNPs relative to sequence length (3 in 1170 bp).

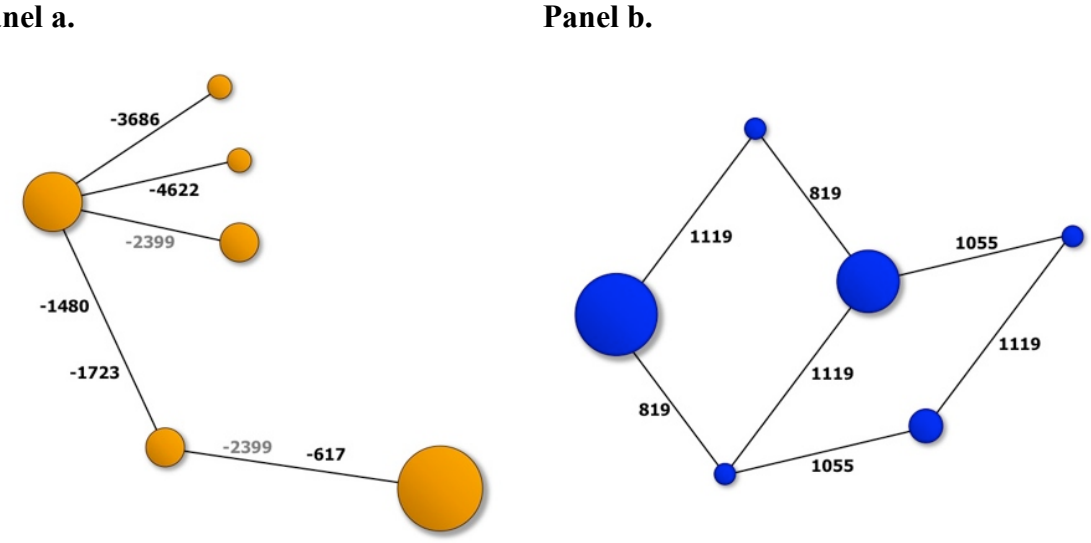


Figure 5.2. Panel a. A median joining (MJ) network of *OXTR* 5' regulatory region sequences from 25 *A. azarai* individuals. Six SNPs were detected in 4936 bp, and defined six haplotypes. Node size is relative to the number of individuals that share a particular sequence. **Panel b.** A median joining (MJ) network of *OXTR* coding region sequences from 25 *A. azarai* individuals. Three SNPs were detected in 1170 bp, and defined six haplotypes. Node size is relative to the number of individuals that share a particular sequence.

OXTR sequence diversity across primates and rodents

In our interspecific comparison of sequence diversity in 40 primate and rodent samples across five regions of the *OXTR* locus, we observed low levels of diversity in coding region sequences (nucleotide diversity: 0.17-0.21) when compared to diversity present in UTR/regulatory and intronic sequences (0.29-0.54). Nevertheless, we were surprised by the amount of variation present across the entire locus among the different taxa, including the substantial numbers of insertion deletions present in the 5' and 3' UTRs (**Table 5.5**).

Our alignment of 40 spliced *OXTR* mRNA sequences (Exons 1 + 2) exhibited 94.8% pairwise identity, with 848 of the 1170 nucleotide positions (72.5%) remaining constant across all taxa, resulting in a nucleotide diversity value of 0.20 for the putative mRNA molecule. The structure of the 7 transmembrane and 3 cytoplasmic domains of the *OXTR* protein was universally consistent in owl monkeys and all other species investigated (**Figure 5.3**). However, the *A. azarai OXTR* mRNA exhibits 4 non-synonymous (d_N) amino acid changes (relative to other primates) that may alter the surface area of the ligand-binding and G-protein binding domains of the mature protein.

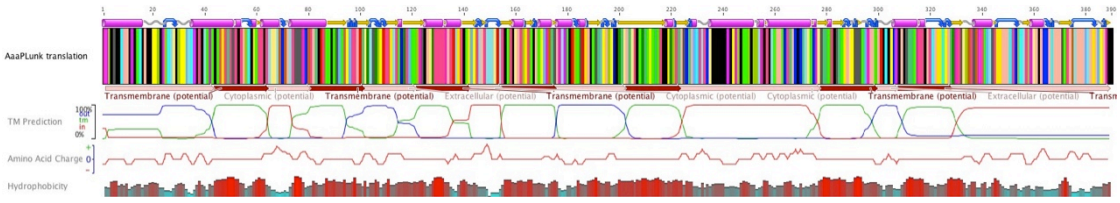


Figure 5.3. A linear representation of *OXTR* mRNA in *A. azarai*. The mRNA sequence runs left to right, 5' to 3', with the MET start codon occurring at position number 1. The seven transmembrane helices, cellular membrane positional prediction, amino acid charge, and molecular hydrophobicity are all represented as parallel tracks beneath the main figure.

Each species possessed lineage-specific *OXTR* coding region nucleotide changes that distinguished it from the other taxa. In particular, the mouse (*Mus musculus*) and rat (*Rattus norvegicus*) both had nucleotides 757-759 (corresponding to codon number 252) completely deleted from their Exon 1 sequences (**Table 5.6**). This codon was present in all primates, suggesting that the presence/absence of the resulting amino acid residue was either a shared derived deletion in rodents, or a shared derived insertion in primates.

Table 5.6. Summary statistics for complete *OXTR* mRNA in 40 primate and rodent samples (26 species and subspecies).

	Spliced *OXTR* mRNA molecule
	1170 bp*
Sequence diversity:	
Pairwise identity of alignment (%)	94.80%
GC content (% non-gap)	64.00%
Nucleotide diversity:	
Nucleotide diversity (avg. over loci)	0.20 +/- 0.10
Summary statistics:	
Usable loci (<5% missing data)	1170
Identical sites	848 (72.5%)
Polymorphic sites	322
Sites with transitions (ts)	224
Sites with transversions (tv)	140
Insertion/deletions (indels)	3**
Selective neutrality:	
Mean # pairwise differences (π)	223.81
Sites with substitutions (S)	314
Tajima's *D* (1000 simul.)	8.09
P (*D* simul < *D* obs)	0.003§
Fu's F_S (1000 simul.)	-3.12
P (sim_F_S ≤ obs_F_S)	0
Amino acid diversity:	
Mean molecular weight (kDa)	42.899
Mean isoelectric point	9.58

* = Indicates width of nucleotide matrix including alignment gaps (to account for insertion/deletion events). The minimum sequence length was 1167 bp (388 amino acids + TGA stop codon) belonging to mm9 and rn4, whereas the *OXTR* mRNA sequences for all primate taxa were 1170 bp (389 amino acids + TGA) in length.
** = The 3 indels reported here constitute a rodent-specific 3 bp deletion of codon number 253, as observed in mm9 and rn4.
§ = *P* value suggests statistical significance.

When comparing 5896 bp of the 5' upstream UTR of *OXTR* in 8 haplorhine primate species, we detected a >250 bp insertion in *Homo sapiens* and a >750 bp insertion in *Pan troglodytes* (**Figure 5.4**). Each of these species-specific insertions was

characterized by large runs of STRs comprised of the dinucleotide motif of "A+T". The 5' UTR sequences of the New World monkeys (*Aotus* and *Callithrix*) were characterized by 10 platyrrhine-specific deletions of 5 bp or greater. One such platyrrhine-specific deletion was 30 bp long, and removed the transcriptional start site (TSS) that putatively functions to initiate transcription in other primate species. Such a molecular event strongly suggests that the platyrrhine TSS has relocated elsewhere, so that *OXTR* transcription can still be initiated. Overall, the presence of such insertion-deletion sequence diversity in the 5' regulatory region of *OXTR* highlights the molecular instability of this non-coding region across primate and rodent genomes.

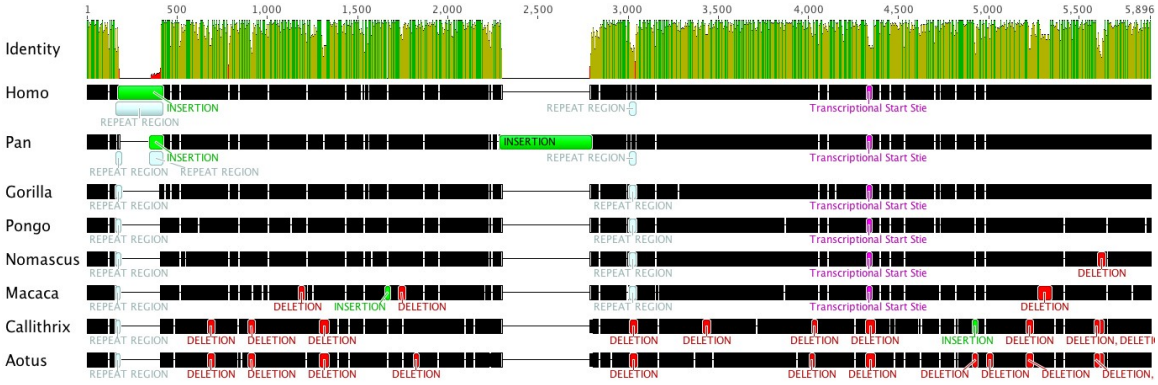


Figure 5.4. A schematic representation of variation in the 5' UTR promoter region of the *OXTR* gene in eight haplorhine primate species. Base numbering runs left to right, 1-5896 bp, 5' to 3', with the MET of Exon 1 occurring immediately following base 5896. Insertions are labeled in *green*, deletions in *red*, repeat regions in *light blue*, and TSS in *purple*. Both *Homo* and *Pan* feature repetitive insertions over 300 bp in size. Both platyrrhine primates (*Callithrix jacchus* and *Aotus azarai*) feature 30 bp deletions which encompass the TSS in other primates, suggesting the initiation of *OXTR* expression in platyrrhines is mediated elsewhere.

To investigate the selective neutrality of *OXTR* in primates and rodents, we calculated Fu's F_S and Tajima's *D* statistics for the 40 sequences in the five different regions of the locus. This analysis revealed Tajima's *D* values ranging from 15.14 (*NS*) in the 5' UTR to 6.54 ($P<0.006$) in Exon 2. Fu's F_S values ranged from 4.1 (*NS*) in the 5'

UTR down to -24.44 (*NS*) in the 39 bp sampling of the intronic region of *OXTR* (**Table 5.4**). The high positive value of Tajima's *D* indicated that an excess of intermediate frequency alleles was present in all regional alignments, and suggested the possibility of balancing selection of heterozygosity (in the form of nucleotide diversity) at the *OXTR* locus in this phylogenetic sampling of primate and rodent genomes. The negative value of Fu's F_S generally confirmed this interpretation.

Phylogenetic analysis of OXTR: Network analysis

The network of *A. azarai* 5' regulatory/promoter sequences of the *OXTR* gene (4936 bp) revealed six distinct haplotypes among the 25 individuals surveyed, which were defined by the six regulatory SNPs (**Figure 5.2a**). The network of *A. azarai OXTR* coding sequences (spliced mRNA alignment) revealed six distinct haplotypes defined by the three exonic SNPs. Three of these haplotypes of these occurred at high frequencies in the population (**Figure 5.2b, Table 5.4**). Upon analyzing the coding sequences from a greater number of owl monkeys within our population (*N*=111) we observed no additional SNPs. As a result, the total number of haplotypes was unchanged in the larger sampling of individuals, and the frequency of each haplotype changed only slightly. Even considering the presence of the three coding SNPs, (all of which are synonymous substitutions), this overall pattern reflected noteworthy nucleotide sequence conservation. Such conservation could be indicative of purifying selection acting upon the specific transcriptional code generated by *OXTR* locus within owl monkeys.

We also created networks for the comparison of 389 amino acids of *OXTR* from primate and rodent species. In these networks, the *Aotus* sequences were clustered with

one another, and separated from other platyrrhine taxa (**Figure 5.5**). Among platyrrhine taxa, *Callicebus* and *Pithecia* clustered together, and both were clearly distinguished from *Lagothrix.* These two clades were both quite distant from *Saimiri*, whereas *Aotus* showed genetic affinities with *Callithrix* and *Saguinus*. In addition, a dramatic number of amino acid changes separated the New World primates from the Old World monkeys and apes, with the latter showing considerably less mutational substructure and shorter internal branch lengths than the platyrrhines.

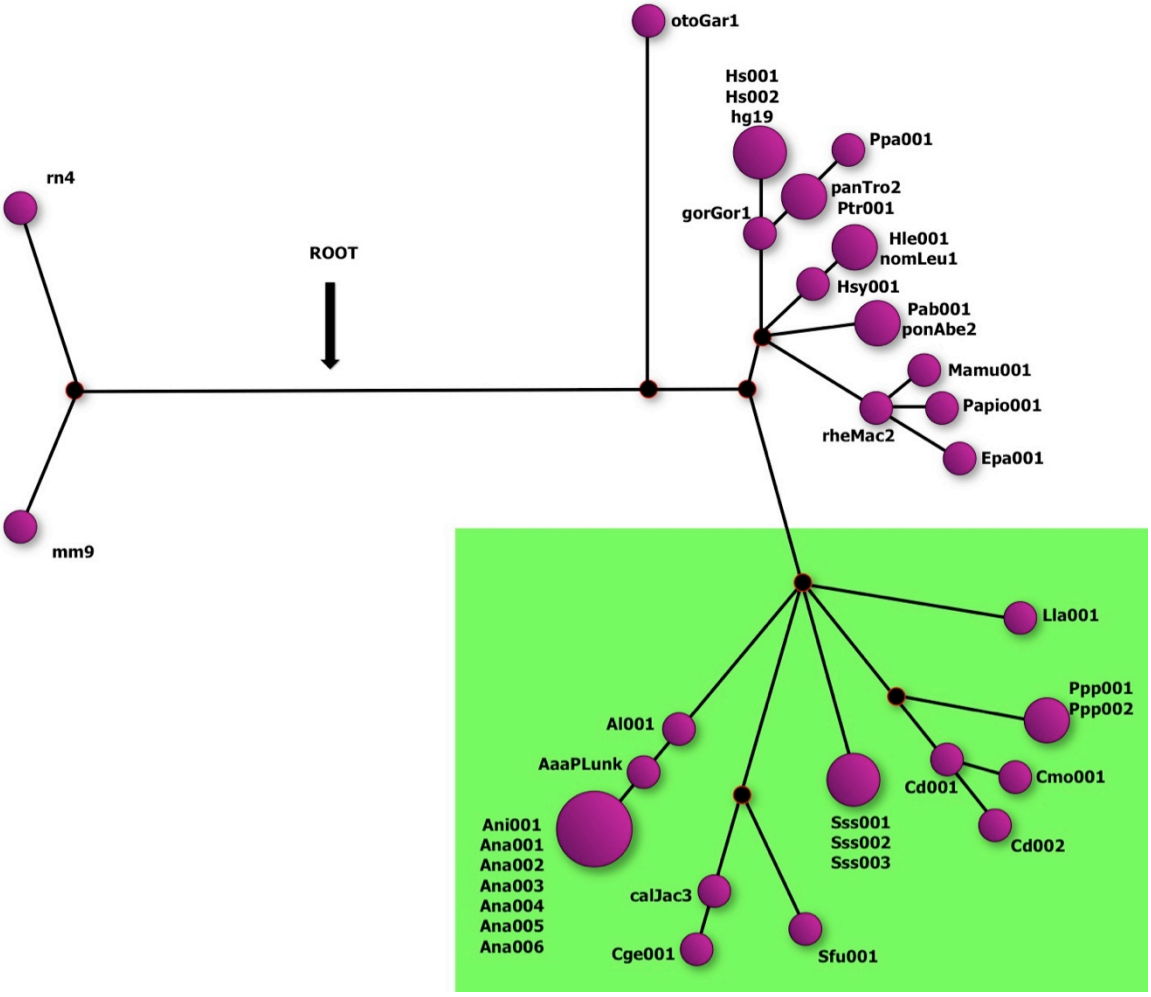


Figure 5.5. A MJ network of 40 *OXTR* amino acid sequences from 26 primate and rodent species. The branches leading to the different clades are proportional to their actual mutation distances. As before, node size is relative to the number of individuals that share a particular sequence motif. The wide range of mutational substructure of amino acid residues is evident in the platyrrhine infraorder, *highlighted* in green.

Phylogenetic analysis of OXTR: Phylogenetic model selection

The AIC model search selected the HKY+G (gamma distribution) model (Hasegawa et al., 1985) with a likelihood score (-ln *L*) of 4411.0792. The BIC and DT model searches chose the TPM2uf+I+G model (invariant sites + gamma), which calculates relationships using six nucleotide states as opposed to only two in the HKY model, and both searches provided likelihood scores of -ln *L*=4414.6732. Because of this discrepancy at the model selection phase, we applied both TPM2uf and HKY models in all subsequent phylogenetic analyses to uncover and evaluate any inconsistencies. Fortunately, we detected no significant changes of phylogenetic arrangements or branch lengths that were specific to the application of one model over the other.

Phylogenetic analysis of OXTR: ML and BI analyses

In agreement with the network analysis, the ML tree representing 1170 bp of the spliced *OXTR* mRNA (**Figure 5.6**) exhibited a similar breakdown of known taxonomic relationships of species based on genetic data (Opazo et al., 2006; Poux et al., 2006; Hodgson et al., 2009; Wildman et al., 2009; Perelman et al., 2011), although the exact phylogenetic relationship that exists between platyrrhine genera is still unresolved. Nevertheless, the majority of phylogenetic relationships among these species remained consistent across the different analyses we performed.

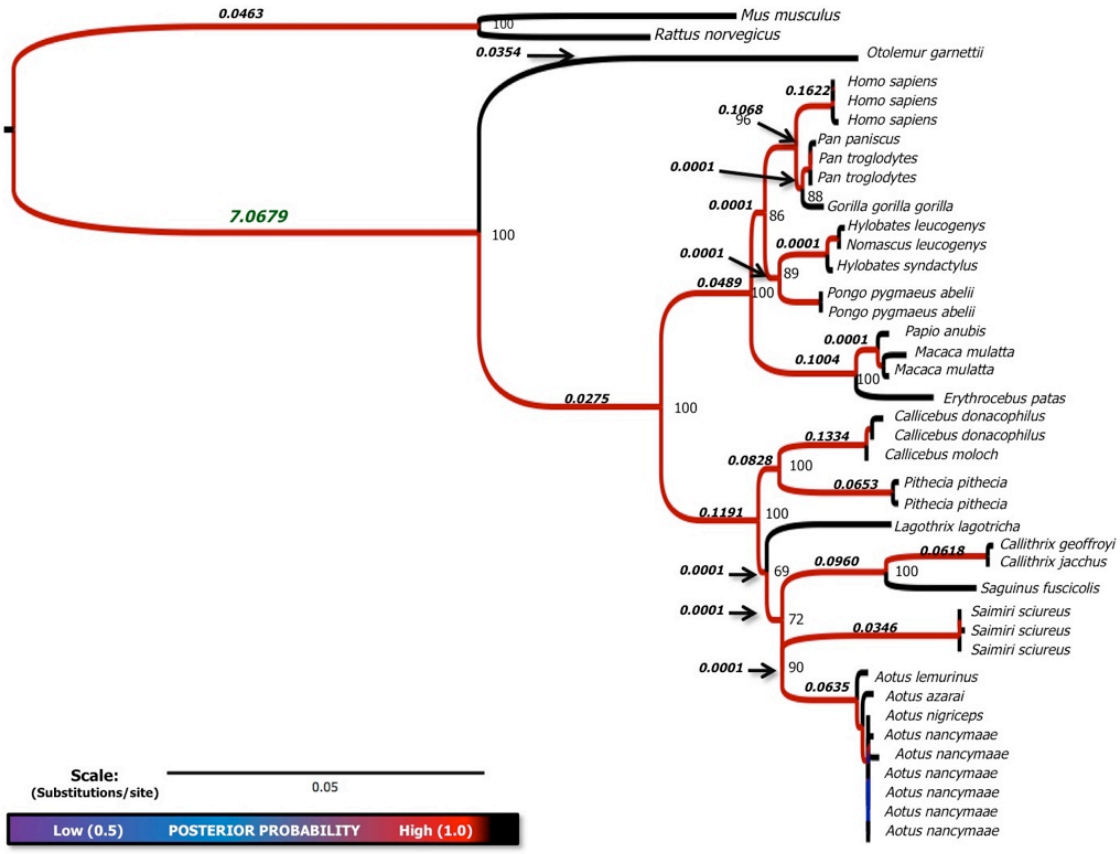


Figure 5.6. A consensus phylogram representing the agreement of our ML and BI phylogenetic arrangements of 40 *OXTR* coding region sequences representing 26 different mammalian species. Branch lengths display the relative number of mutational differences detected for each different taxonomic group. The d_N/d_S values are shown above each branch, whereas ML bootstrap values (of 10,000 replicates) appear below them. For the BI analysis, two MCMC chains were run in parallel across 500,000 generations, with trees sampled every 100 generations. A total of 5,001 trees were retained and sampled for a 50% majority consensus tree, after a 1,000-tree burn-in. Posterior probabilities of the phylogenetic arrangement are displayed as a color gradient across taxa.

Phylogenetic analysis of OXTR: Analysis of adaptive evolution

Our results indicated differential rates of evolution of the *OXTR* locus in different primate lineages. Using the ML tree calculations, we predicted the relative rates of *OXTR* amino acid changes along the different phylogenetic branches in disparate mammalian taxa to detect any signatures of adaptive evolution. The resulting d_N/d_S ratio values for the majority were exceptionally low (d_N/d_S=0.0001) along branches leading to different primate genera-based clades, suggesting the forces of purifying selection are

shaping diversity at the *OXTR* locus (Yang, 2007) (**Figure 5.6, Figure 5.7**). We noted stronger positive selection signals (d_N>d_S) on branches leading to the taxonomic outgroups, particularly on the branch distinguishing the primate and rodent orders (d_N/d_S=7.0679). These signals were further illustrated when comparing the pure d_N tree to the pure d_S arrangement (**Figure 5.7**). The LRT indicated that the M1 (variable) model was more likely (-ln *L*=4292.5301) than M0 (homogenous d_N/d_S=0.0546 for the entire tree) model (-ln *L*=4258.3135), although this result was non-significant (*P*>0.8381). Therefore, we decided that individual branch values would more accurately describe our data with regard to the adaptive evolution of *OXTR* and directional selection in primates.

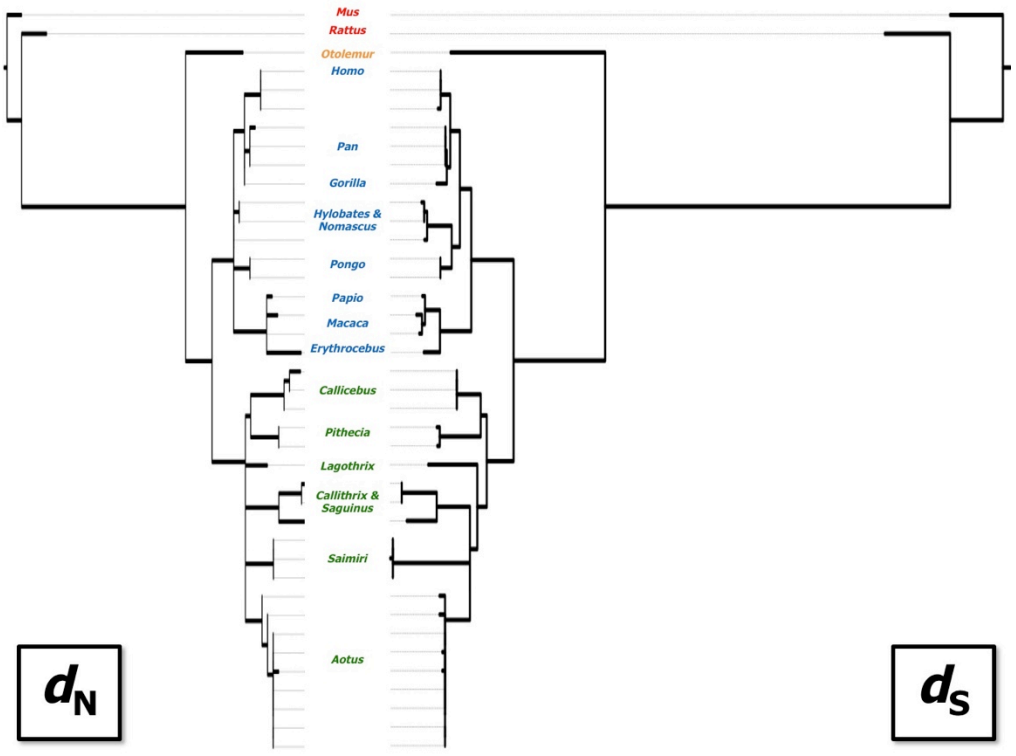


Figure 5.7. Oppositional phylogram representing the different phylogenetic patterns and branch lengths based on *OXTR* coding sequence variation and exhibited by d_N and d_S trees.

Phylogenetic analysis of OXTR: Bayesian coalescent estimation

The BEAST analysis produced a highly resolved phylogenetic chronogram of primate and mammalian species with *OXTR* coalescent time points and mutation rate estimates for each branch of the tree. Its topology presented a taxonomic arrangement of these species that was identical to those produced by the ML and median network analyses (**Figure 5.8**). In addition, through this Bayesian analysis, we were able to visualize the changes in mutation rate that occurred along different evolutionary branches, as well as generate age estimations at each phylogenetic node. The TMRCA estimations and their associated error ranges were consistent with those in other recent studies of molecular data (Hodgson et al. 2009). These conditions strengthened our confidence in the evolutionary scenario for the evolution? of *OXTR* among the *A. azarai* and platyrrhine primates that we propose below.

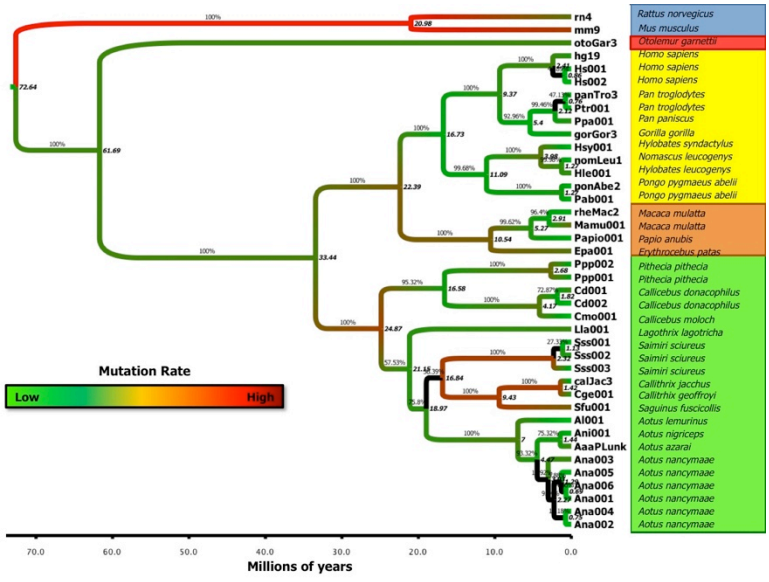


Figure 5.8. A chronogram depicting the Bayesian coalescence age estimation of *OXTR* coding region sequences, as calculated in BEAST v1.5.4. Age estimates for each node are italicized and positioned to the right of them. Posterior probabilities are indicated above each branch. Mutation rate of *OXTR* along the different evolutionary lineages is displayed as a red-green color gradient across taxa. The different taxa compared here are displayed on colored backgrounds corresponding to their phylogenetic placement. *Blue* rodents, *red* prosimians, *yellow* apes, *orange* Old World monkeys, *green* New World monkeys.

5.4 Discussion

OXTR evolution in A. azarai

The individuals from our owl monkey population exhibit very little coding region diversity for the *OXTR* gene. In addition, the overall sequence diversity present in the 5' UTR of owl monkeys is rather limited, with only six (6) SNPs present in 4936 nucleotides. For the entire *OXTR* locus, we observed a total of only 96 single nucleotide polymorphic events in the 313,600 nucleotides (diploid) surveyed in 25 Azara's owl monkeys.

One scenario that would explain these data is that purifying (negative) selection is taking place across the entire *OXTR* locus (UTRs, exons, and intron) in *A. azarai*. Negative selection occurring at the regulatory region could serve the purpose of maintaining a specific level of functional gene expression, whereas conservation at the coding region could preserve particularly optimized protein structures. Owl monkey 5' regulatory regions display 177 TFBSs, whereas other species range from 171 in *Pongo,* to 213 in *Pan* (**Figure 5.9**). As seen in other platyrrhines, the transcriptional start site of owl monkey *OXTR* is absent, due to a 30 bp deletion) suggesting that effective *OXTR* transcription must be initiated by other means, either proximately in the locus, or elsewhere in the genome. The presence and absence of such transcriptionally relevant elements could act to enhance or dampen *OXTR* expression in owl monkeys.

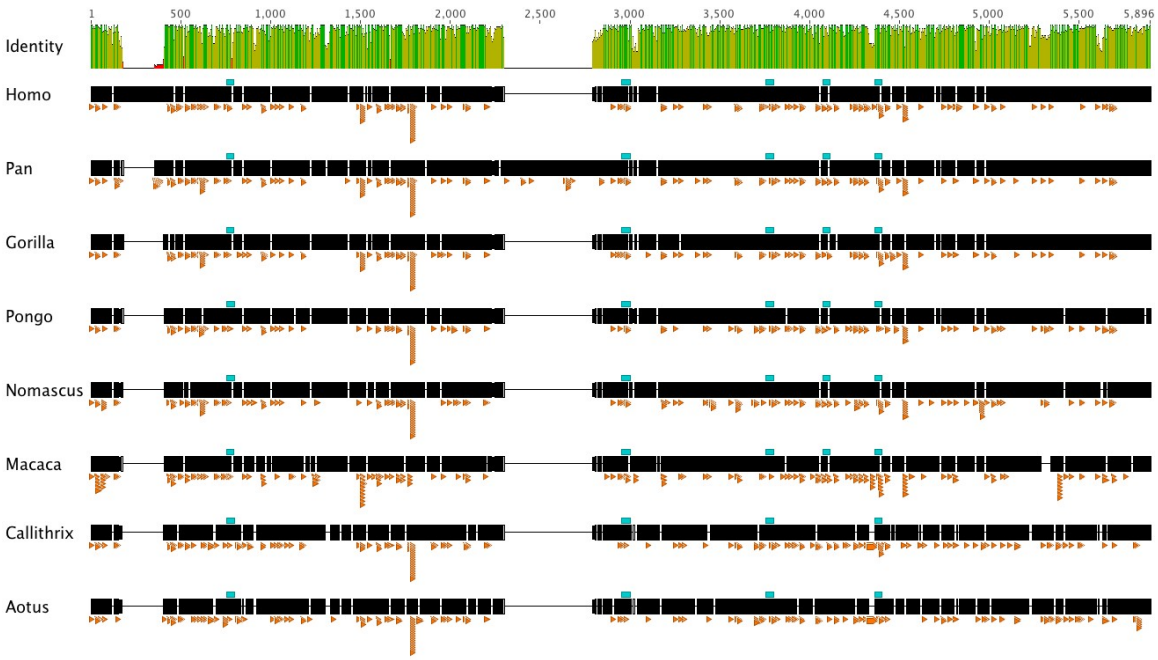


Figure 5.9. A diagram of the 5' UTR promoter region of the *OXTR* gene in eight haplorhine primate species, highlighting putative locations of transcription factor binding sites (TFBS) with lengths of 6 bp or more. Base numbering runs left to right, 1-5896 bp, 5' to 3', with the MET of Exon 1 occurring immediately following base 5896. Transcription factor sequence recognition sites are labeled in *orange* beneath the sequences, whereas poly-A signals are noted above the sequences in *blue.*

The non-synonymous coding region mutations observed in the *OXTR* coding sequences in *A. azarai* and other taxa imply that such structural alterations have taken place at the ligand and G-protein binding domains in the mature *OXTR* protein (**Figure 5.10, Figure 5.11**). While these types of non-synonymous amino acid substitutions and their location in the molecular structures of *OXTR* vary across species, they are fixed within the *A. azarai* population. Thus, it is likely that such changes could alter the functional activity of OT receptor neurons, and could impact the manifestation of the sociobehavioral phenotype of long-term partner preference.

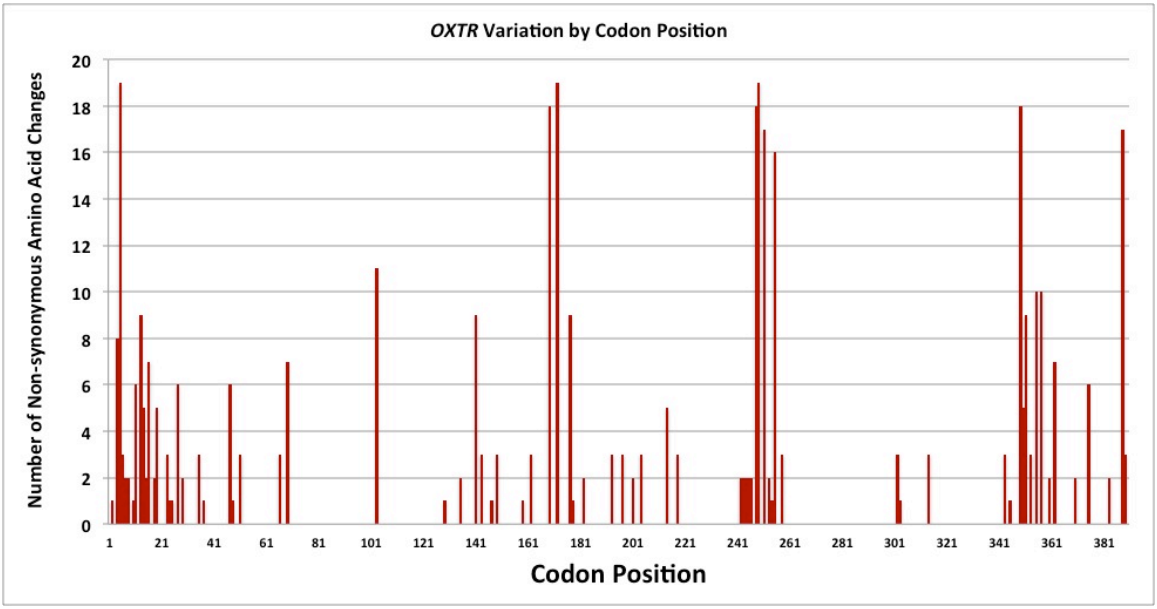


Figure 5.10. A stacked histogram displaying the number of non-synonymous amino acid changes of the *OXTR* coding sequence in 40 different species, relative to a translated consensus DNA sequence (assessed by plurality, [50% majority rule]). The interspecific clustering of these changes would indicate the localization of amino acid variation at putatively functional portions of the mature *OXTR* mRNA. The *x*-axis represents the location of those changes along the first exon of *OXTR*, with the ATG situated at the graph's origin (0).

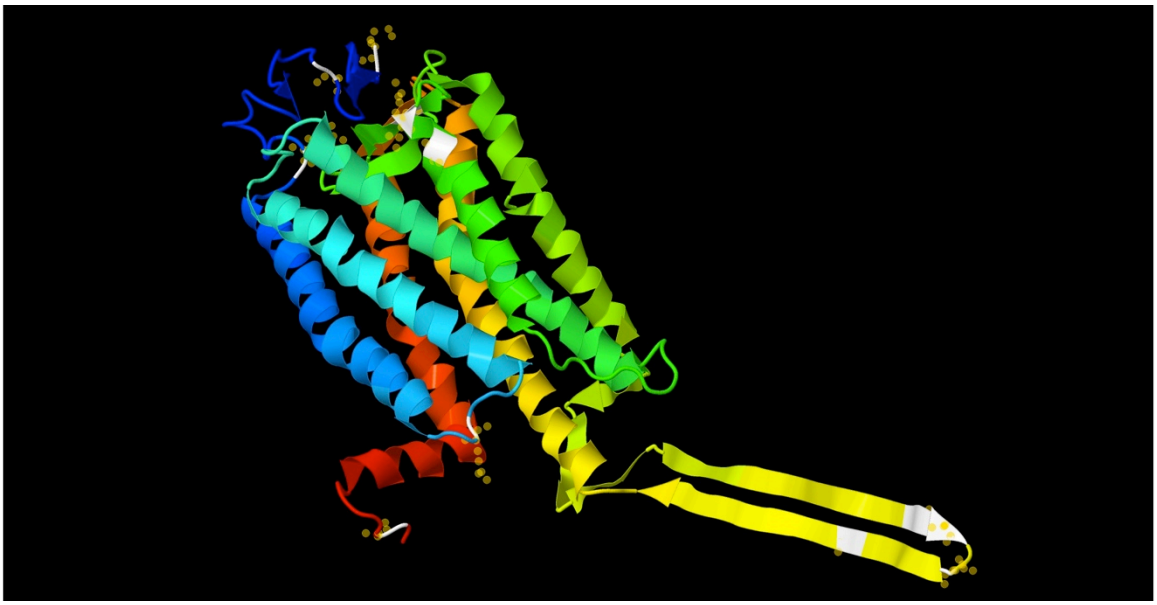


Figure 5.11. A two-dimensional projection of non-synonymous amino acid changes specific to *A. azarai OXTR* coding sequence, as derived from the codon position variance histogram (**Figure 5.7**). White stripes and yellow translucent circles represent amino acids positions where *A. azarai OXTR* mRNA structurally differs from *Homo sapiens OXTR* mRNA.

OXTR evolution in platyrrhines

Despite the polygenic characteristics of the OT pathway and the high levels of sequence conservation among Azara's owl monkeys, our findings suggest that *OXTR* proteins have undergone a considerable amount of evolutionary change across primate taxa, as exhibited by the wide range of variation in the *OXTR* coding and regulatory regions (**Figure 5.4, Figure 5.8, Figure 5.9**).

Platyrrhine species have accumulated many synonymous and non-synonymous mutations in the coding sequence of *OXTR* when compared to Old World monkeys and hominoid primates. Each of the seven platyrrhine genera examined in this study (*Aotus*, *Lagothrix, Saguinus*, *Callithrix*, *Callicebus*, *Pithecia* and *Saimiri*) exhibits a number of lineage-specific mutations that distinguishes their *OXTR* coding sequences from those of any other taxon being investigated. Moreover, *Callithrix*, *Saimiri* and *Saguinus* genera of New World monkeys possess nearly twice as many amino acid changes (from 40 sequence consensus) than do other platyrrhines, and feature a faster rate of mutational accumulation according to the BEAST analysis (**Figure 5.8**).

Amino acid changes have also accumulated in different primate species in a lineage-specific manner. The distribution of amino acid changes is notably diverse within the platyrrhine infraorder, compared to the localization of non-synonymous changes observed in other taxa (**Figure 5.12**). However, when compared to the consensus of the 40 sequences examined, hominoid species exhibit, on average, more non-synonymous mutations in *OXTR* mRNA sequences than the other tax (with the exception of the two rodents and bushbaby (**Figure 5.13**). Yet, platyrrhines also exhibit the shared derived possession of six amino acid changes in their *OXTR* mRNA that

separate their clade from that of other primates (**Figure 5.5, Figure 5.6**). These changes on the receptor could have been a driver of, or a response to, the novel form of the OT neuropeptide that was recently discovered in New World monkeys (Lee et al., 2011).

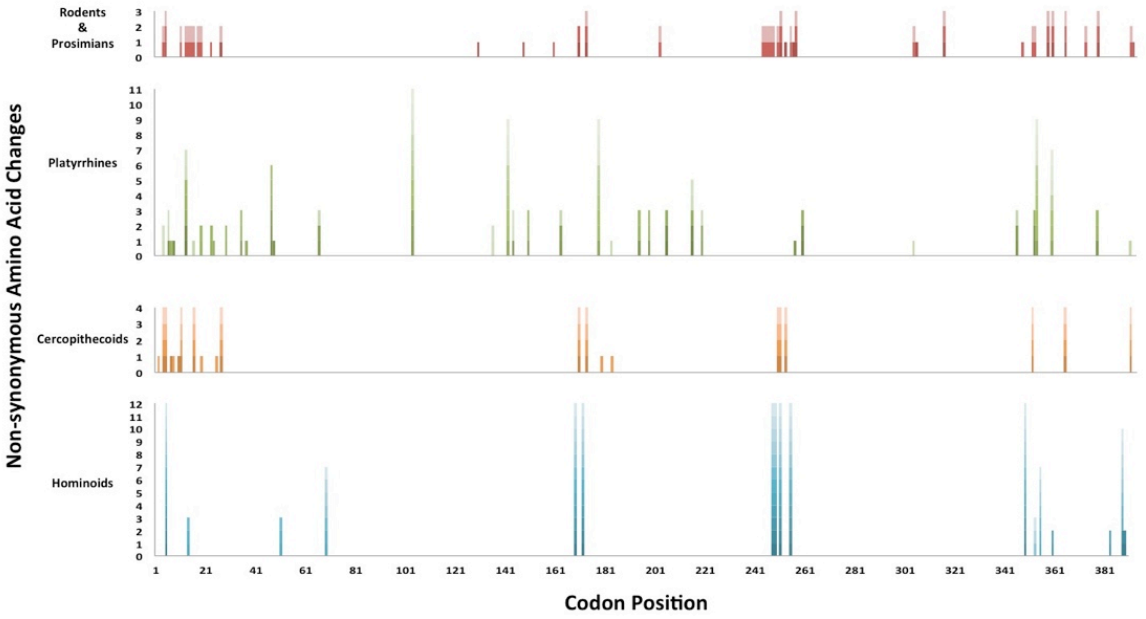


Figure 5.12. Stacked histograms displaying the number of non-synonymous amino acid changes of the *OXTR* coding sequence (*x*-axis represents codon numbers 1-390) in four taxonomic categories: rodents & prosimians, platyrrhines, cercopithecoids, and hominoids. Amino acid changes are relative to a translated consensus DNA sequence derived from the alignment of 40 *OXTR* mRNA sequences (assessed by plurality, [50% majority rule]). The interspecific clustering of these changes would indicate the localization of amino acid variation at putatively functional portions of the mature *OXTR* mRNA. The distribution of amino acid changes is notably diverse within the platyrrhine infraorder, compared to the localization of non-synonymous changes observed in other taxa.

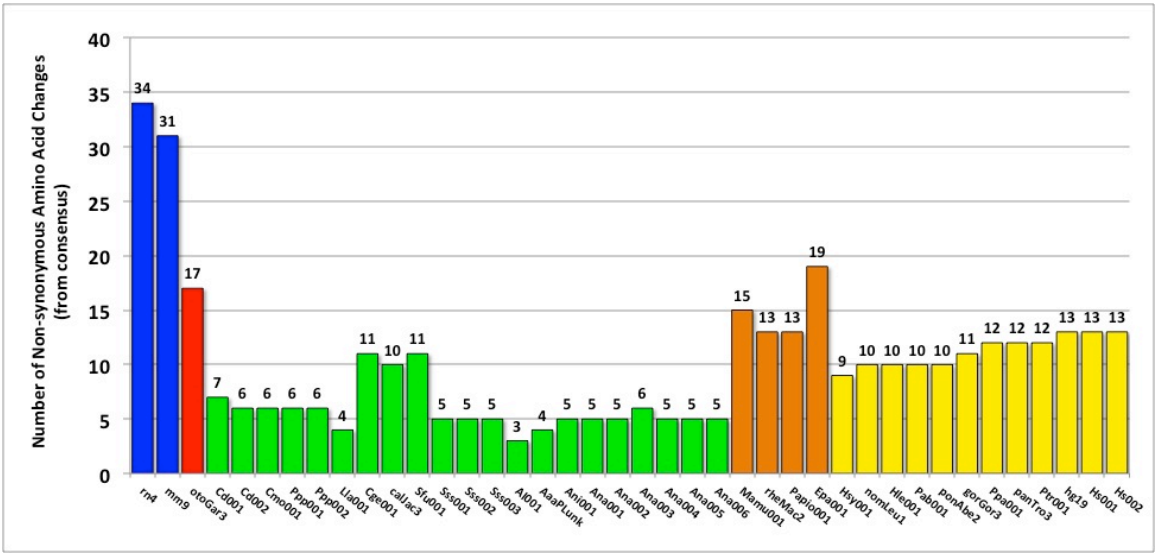


Figure 5.13. The frequency of non-synonymous amino acid changes in the *OXTR* gene, as calculated from the consensus translation (by plurality, [50% majority rule]) of 40 rodent and primate sequences. The different taxa compared here are displayed as colored bars corresponding to general taxonomic categories. *Blue* rodents, *red* prosimians, *yellow* apes, *orange* Old World monkeys, *green* New World monkeys.

Although the *OXTR* coding diversity observed in platyrrhines is undoubtedly a result of the colonization of the New World by proto-platyrrhines some 30 million years ago (Poux et al., 2006; Hodgson et al., 2009), it may also reflect the consequences of selection acting upon this locus, especially given the enrichment of monogamous social systems among New World species (Di Fiore and Rendall, 1994; Schwindt et al., 2004; Rendall and Di Fiore, 2007).

OXTR evolution in primates

That selection may be at work at this locus across the Order Primates is indicated by the statistical analysis of *OXTR* coding sequences. The high positive value of Tajima's *D* and negative value of Fu's F_s reveals an excess of intermediate frequency alleles, which, in turn, suggests that the observed variation has been shaped by balancing selection. Thus, despite high levels of intraspecific coding conservation in 25 *A. azarai*

individuals, the *OXTR* gene exhibits considerable nucleotide and codon diversity among closely related primate species, and the presence of large indels and derived clusters of STRs in the *OXTR* 5' regulatory region present another tier of variation altogether.

Humans and chimpanzees have each experienced unique insertions of repetitive sequences roughly 4500 (*Homo sapiens*) and 2500 (*Pan troglodytes*) bp upstream of the *OXTR* start site. Such promoter insertions can greatly affect the efficiency of transcription of the related gene. It is therefore not unreasonable to hypothesize that different lineages of primates transcribe *OXTR* at different rates or in the presence of different transcriptional co-factors.

While searching for more information on the transcriptional control of expression of *OXTR* in primates, we compared our aligned sequences with the genomes of other primate species on UCSC GB (Kent et al., 2002; Karolchik et al., 2008). Our genomic comparison of structural variation at the chromosomal level revealed that, despite SNPs and insertion-deletion events, major chromosomal rearrangements have not reshuffled the regulatory regions upstream of the gene in primates (**Figure 5.4, Figure 5.9**). This greatly distinguishes *OXTR*'s evolutionary history from that of *AVPR1A* (see Chapter 4), which experienced a chromosomal rearrangement in platyrrhines.

However, when surveying this region using the DGV (Iafrate et al., 2004) track on UCSC GB, we noted that, for many primate taxa, CNVs exist in large sections of the entire region (Gregory et al., 2009). Recent genomic studies have noted that simple variation (STRs), similar to those upstream of the *OXTR* locus in humans and chimps, may precipitate more mutations, both base substitutions and structural changes (Conrad et al., 2010). Repetitive sequences that are enriched in promoter regions are also enriched

in CNV breakpoints, suggesting that the same properties that enable regulation of transcription may also be mildly mutagenic for the formation of CNVs, and as a consequence, CNVs may influence the evolution of gene regulation (Conrad et al., 2010).

Role of OXTR in the evolution of primate sociality

Ecological conditions may be a strong selective influence on the evolution of primate social behavior, and molecular sequence changes could serve as a means of adapting to these ecological challenges. Across the Order, primates possess notable levels of interspecific variation in their *OXTR* coding regions. Regulatory regions provide additional locations for mutations to accumulate, and, in some cases, could facilitate quick adaptive responses (via gene transcriptional regulation) to environmental change. Therefore, a certain level of plasticity in these regions may be tolerated within populations.

Many molecular mechanisms can reshape genetic sequences and potentially alter gene dosage and function. CNVs and STRs are present throughout the *OXTR* region in mammals (Kent et al., 2002; Karolchik et al., 2008), and could facilitate changes in gene expression in different species or even individuals (Perry et al., 2007; Gökçümen and Lee, 2009; Conrad et al., 2010; Park et al., 2010). Therefore, the careful investigation of both coding and non-coding regions for signatures of selection is essential for elucidating the evolution of partner preference and pair bonding in extant primate species. Moreover, as functional variants at *AVPR1A* and *OXTR* have been reported in humans (Takayanagi et al., 2005; Hammock & Young, 2006; Egashira et al., 2007; Grippo et al., 2008, 2009; Sebat et al., 2007), our study has direct implications for understanding inter-

individual variation in these genes in humans by using owl monkey populations and social groups as comparative models.

In conclusion, we have characterized the molecular features of *OXTR* in *A. azarai* and other platyrrhines, and demonstrated that its sequence differs from that seen at this locus in other primate species. These data provide new clues about the possible basis of pair bonding in New World species, and may help to explain the sporadic appearance of monogamy in this infraorder. More specifically, despite a common molecular origin, we argue that the AVP and OT pathways have evolved in markedly different ways, due in part to their chromosomal locations and their relative proximity to regions of molecular instability. This study reinforces the notion that primates have experienced significant neurogenetic variation during their evolution, and suggests that monogamy has arisen multiple times in the primate order through different molecular mechanisms.

Chapter 6: Expression

Summary

AVPR1A and *OXTR* are widely expressed in the brains of mammals, and both genes are considered to be key loci for the regulation of social behavior. Studies in several animal models have demonstrated that variation in and around these genes directly influences the manifestation of different socio-behavioral phenotypes. While amino acid substitutions in the coding regions of both loci undoubtedly affect the final protein structure of *AVPR1A* and *OXTR*, mounting evidence has implicated variation in the 5'-flanking regions of both loci in variable gene expression and social behavior.

Primate species, including humans, exhibit a wide range of complex social behaviors, yet the extent to which *AVPR1A* and *OXTR* expression varies in the brains of those primate taxa remains unknown. To better understand this range of transcriptional variation, we undertook promoter-reporter analyses of both genes in six primate species that exhibit notably different forms of social behaviors and mating systems. After comparative sequence analyses, we isolated the regulatory regions of *AVPR1A* and *OXTR* using long-range PCR (LR-PCR), and inserted them into dual-luciferase promoter reporter vectors. We transfected these artificial bacterial vectors into human neuroblastoma cells, and initiated their expression to quantify and compare potential transcriptional differences of these genes in different primate species. Our results suggest that different sequence motifs lead to significant variation in *AVPR1A* or *OXTR* mRNA production across primate taxa, and that such species-level sequence variation might help explain the diversity and intensity of pro-social behaviors in different primates.

6.1 Introduction

The neuropeptides AVP and OT have been implicated in the manifestation of social bonding, and in the regulation of a variety of socially relevant behaviors in mammals (Carter et al., 2008). Pharmacological experiments in rodents have demonstrated roles for both AVP and OT in social learning, memory, aggression, and affiliative behaviors (Carter, 1998; Hammock & Young, 2004, 2005; Young et al., 2005). Specifically, AVP and OT act by binding to ligand-specific GPCR proteins, *AVPR1A* and *OXTR*, to transfer signals at neuronal synapses and across cellular membranes. In non-human mammals, the variable expression of the *AVPR1A* and *OXTR* genes is purported to alter the density and distribution of those receptors in the brain, which in turn, has been suggested to influence in changes in social bonding, parental rearing behaviors, social recognition, and memory.

Non-human primates display a wide range of pro-social behaviors, and the *AVPR1A* and *OXTR* genes of different primate species exhibit a large amount of sequence variation (see Chapters 4 and 5). This sequence variation exists as amino-acid coding changes, which potentially alter the structure of the transcribed mRNA and the translated GPCR proteins in different primate species. In addition, primates also exhibit tremendous variation in the regulatory regions associated with the *AVPR1A* and *OXTR* genes, raising the probability that different taxa express these genes at different rates (Hong et al., 2009). Thus, it is possible to envision a scenario wherein increases or decreases in *AVPR1A* and *OXTR* expression of a particular primate could putatively influence the manifestation and intensity of pro-social behaviors of a particular taxonomic group.

To test this hypothesis, we sequenced the regulatory regions of *AVPR1A* and *OXTR* in six primate species (*Aotus azarai*, *Callithrix geoffroyi*, *Macaca mulatta*, *Pongo pygmaeus abelii*, *Pan troglodytes*, *Homo sapiens*), which exhibit a large range of social and mating behaviors. After making a comparative sequence analysis, we re-isolated and amplified the regulatory regions using long-range polymerase chain reaction (LR-PCR), and inserted the fragments into dual-luciferase promoter reporter vectors to drive synthetic firefly luciferase genes. We transfected these artificial bacterial vectors into human neuroblastoma cells, and initiated their expression to quantify and compare potential transcriptional differences of these genes in different primate species.

6.2 Methods

The overall workflow of this study is presented in **Figure 6.1**.

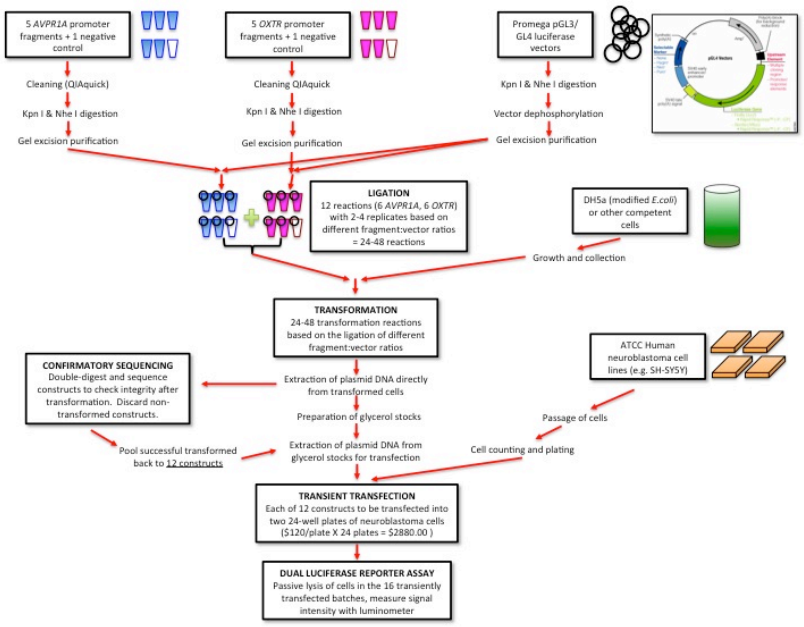


Figure 6.1. Experimental workflow depicting the creation of long-range PCR templates, vector digestion and ligation, *E. coli* transformation, neuroblastoma cell transfection and luciferase assay luminescence reading.

Samples

We assembled a panel of six individuals representing six species of primates (*Aotus azarai*, *Callithrix geoffroyi*, *Macaca mulatta*, *Pongo pygmaeus abelii*, *Pan troglodytes* and *Homo sapiens*) for the functional quantification of expression associated with interspecific variation of *AVPR1A* and *OXTR* regulatory sequences (**Table 6.1**). The *A. azarai* individual in this panel (Aaa021) possesses the most frequent haplotype of the *AVPR1A* and *OXTR* regulatory regions that we observed in the target wild study population (n=118) (Chapters 4 and 5). The human DNA (Hs001) was obtained from buccal cells extracted by, and belonging to, the author (P. Babb). In addition, we purchased comparative samples for four other primate taxa from the NIA and IPBIR cell line collections, curated by the Coriell Institute of Biomedical Research (Camden, NJ).

Table 6.1. Samples investigated for expression variation at the *AVPR1A* and *OXTR* loci.

ID	Species	Common Name	Sex	Social Organization	Locale
Aaa021	*Aotus azarai*	Azara's Owl Monkey	M	Monogamous	Core area Formosa, AR
Cge001	*Callitrhix geoffroyi*	White-fronted Marmoset	F	Monogamous/Polyandrous	Coriell Institute (IPBIR: PR01094)
Mamu001	*Macaca mulatta*	Rhesus Macaque	F	Polygynous	Coriell Institute (NIA: NG06249)
Pab001	*Pongo pygmaeus abelii*	Sumatran Orangutan	M	Solitary	Coriell Institute (IPBIR: PR00253)
Ptr002	*Pan troglodytes*	Chimpanzee	F	Multimale-Multifemale	Coriell Institute (IPBIR: PR00605)
Hs001	*Homo sapiens*	Human	M	Multimale-Multifemale/ Monogamous(?)	Self (P.Babb)

Note: Core study area is located in Formosa Province, Argentina (Lat. = 25º, 59.4 min South; Long. = 58º, 11.0 min West).

We also conducted searches of the publically available genomes accessible through the UCSC Genome Browser and the ENSEMBL browser for the extraction of

annotated primate *AVPR1A* and *OXTR* sequences (Kent et al., 2002; Karolchik et al., 2008).

Long-Range PCR and Sequencing

To investigate interspecific variation at the regulatory regions and putative promoters of the *AVPR1A* and *OXTR* genes, we targeted roughly six kilobases (kb) of adjacent non-coding sequence at the 5' end of each gene (Inoue et al., 1994; Thibonnier et al., 1996; Kent et al., 2002). For the *AVPR1A* locus, this ~5.2 kb region included: 3.2 kb of 5' flanking sequence upstream of the transcriptional start site (TSS) and the entire 1.5 kb of the 5' untranslated region (UTR), and 622 bp of the first coding exon (**Figure 6.2**). For the *OXTR* locus, this ~5.9 kb region included: 3.4 kb of 5' flanking sequence upstream of the transcriptional start site (TSS) and the entire 1.5 kb of the 5' untranslated region (UTR), the first coding exon (922 bp), and 132 bp of the 5' end of the intron (**Figure 6.3**). For both loci, we focused specifically on orthologous sequences (>65% pairwise similarity) across taxa, and therefore excluded regions representing taxon-specific chromosomal crossovers and structural variants.

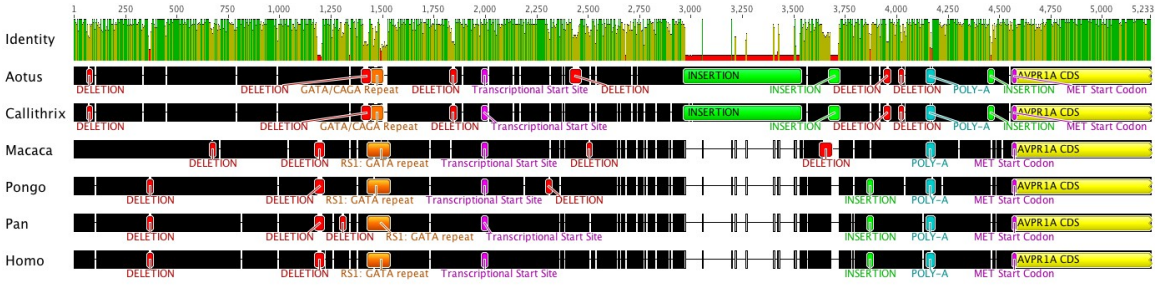


Figure 6.2. *AVPR1A* regions inserted into promoter-reporter plasmid vectors.

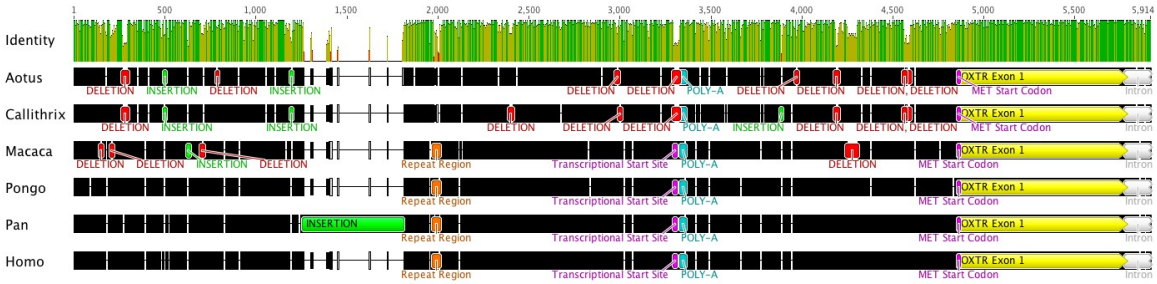


Figure 6.3. *OXTR* regions inserted into promoter-reporter plasmid vectors.

In order to directly amplify these regions, we generated >20 kb interspecific *AVPR1A* and *OXTR* genic contig alignments from genomic sequence files retrieved from UCSC GB, and ENSEMBL using Sequencher, v4.9 (Gene Codes) and Geneious Pro v5.5.5 (Drummond et al., 2010). From these assemblies, we designed four unique *AVPR1A* and *OXTR* primers to bind in regions exhibiting high levels of sequence conservation across all taxa, using the oligo software programs NetPrimer (Premier BioSoft) and Primer3 (Rozen and Skaletsky, 2000). For each gene in each sample, these primers amplified a long-range fragment (5-6 kb in size) of orthologous regulatory promoter sequence present in the different primate species under investigation. We then designed "tailed" primers featuring 5' tags containing restriction site(s) and a non-complementary tail for use in enzyme-based vector ligation (described below). In addition, we utilized alternate primers situated within amplicons to facilitate confirmatory sequencing (see Chapters 4 and 5) (**Table 6.2**).

Table 6.2. Primer sequences designed and optimized for *AVPR1A* and *OXTR* LR-PCR amplification.

Genetic Region	Primer Name	Strand	Oligonucleotide Sequence (5'-3')**	Restriction Enzyme
AVPR1A 5' Upstream Region	AVPR1A-LR04-F-tail3	Forward	GCAC CTCGAG AAG TGG CTC AAC AGT CAA AG	Xho I
	AVPR1A-pcr3-R-tail3	Reverse	GCACAGATCT GTC ATC CAG GTC ACG TAG GC	Bgl II
OXTR 5' Upstream Region	OXTR-LR03-F-tail2	Reverse	GCACGGCCTAACTGGCC TGA CCT CCC TCA GCA AGA AG	Sfi I
	OXTR-Ex1-R3b-tail2	Forward	GCACAAGCTT GAC TCT GTG GGA TTT CAA AC	Hind III

**Non-complimentary primer tails are marked with a grey background, enzymatic cut sites are underlined, and the binding primer sequence is denoted by boldface type.

Long-range polymerase chain reaction (LR-PCR) cycling parameters were optimized for each primer pair using a Touchgene Gradient thermocycler (Techne), and all subsequent reactions were amplified in GeneAmp 9700 thermocyclers (ABI). We used the Roche Expand Long-Range PCR dNTPack (Roche Applied Science) amplification kit, which takes advantage of two polymerase enzymes to isolate and proofread the appropriate genetic targets. Cycling parameters and LR-PCR recipes are detailed in **Table 6.3**. Amplified products were visualized on 0.7% TBE SeaKem agarose (Lonza) via gel electrophoresis. Gels were run for three hours at 85 V and 100 mA.

For confirmatory sequencing, amplicons were purified through SAP/Exo I digestion (New England BioLabs), and then cycle-sequenced using BigDye^{TM} v3.1 (ABI). Excess dye terminators were removed with the BigDye XTerminator^{TM} purification kit (ABI), and DNA sequences were read on a 3130xl Gene Analyzer (ABI).

Table 6.3. Long-Range PCR cycling parameters.

Genetic Region	Initial Denature	First segment (10 cycles)			Second segment (30 cycles)			Final Extension
		Denature	Anneal	Extend	Denature	Anneal	Extend	
***AVPR1A* 5' region PCR temp:**	92°C	92°C	50°C	68°C	92°C	50°C	68°C	68°C
Cycle time (minutes):	2:00	0:10	0:30	6:00	0:10	0:30	6:00 + 0:20cycle	7:00
***OXTR* 5' region PCR temp:**	92°C	92°C	49°C	58°C	92°C	49°C	58°C	58°C
Cycle time (minutes):	2:00	0:10	0:30	6:00	0:10	0:30	6:00 + 0:20cycle	7:00

Note: Recipe components for all 50 µL LR-PCR reactions: 13.75 µL ddH_20, 10 µL 5X Expand LR Buffer without $MgCl_2$, 4 µL 25mM $MgCl_2$, 2.5 µL 10 mM 'dNTPack' dNTP pre-mix, 5 µL forward primer (10 pmol/ µL), 5 µL reverse primer (10 pmol/ µL), 5 µL DMSO, 0.75 µL Expand LR Enzyme mix (*taq* and *pfo*), 4 µL DNA (5 ng/µL). All reactions underwent a final infinite hold at 4°C.

We assessed read quality for each sequence using Sequencing Analysis v5.4 (ABI) and aligned them using Sequencher and Geneious. Using this approach, we confirmed our ability to reliably isolate 5.2 kb of orthologous *AVPR1A* 5' regulatory sequence and 5.9 kb of orthologous *OXTR* 5' regulatory sequence for each sample.

Sequence alignments

For each sample, we assembled sequences for the regulatory regions of *AVPR1A* and *OXTR* by aligning the overlapping forward and reverse fragments using strict (95%) agreement thresholds. To investigate the variation across the *AVPR1A* and *OXTR* regulatory regions, we constructed two alignment matrices of the six samples to examine the nucleotide diversity across the region for all individuals. These matrices were 5233 bp long for the *AVPR1A* regulatory region, and 5914 bp long for the *OXTR* regulatory region.

Using this collection of sequences, we searched for any polymorphic mutations

and putative TFBS in the upstream promoters of both genes. This analysis was conducted using the MatInspector software package (GenoMatix Software, GmbH), with parameters being set for transcription factors present in mammalian neuronal cellular tissues as derived from human and mouse consensus databases. This work was confirmed using the transcription factor prediction software plugin for Geneious v5.5.5 (Drummond et al., 2010).

Allele identification, gel slicing & purification

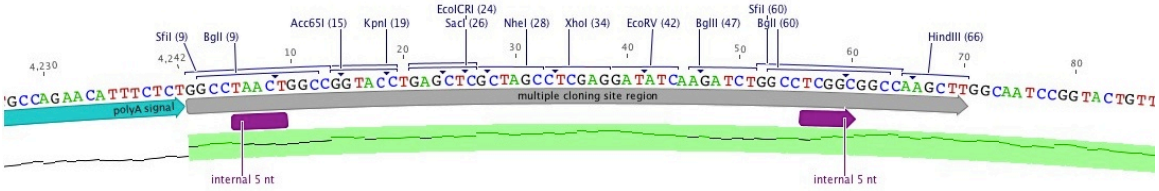
Following successful LR-PCR of the six *AVPR1A* and six *OXTR* fragments, we re-isolated the regions using "tailed" primers to facilitate plasmid construction. These primers were designed to possess a 5' non-complementary tail engineered to contain Sfi I and HindIII (*AVPR1A*) and Xho I and Bgl II (*OXTR*) restriction enzyme cut-sites to assist the ligation of the regions to the vector. We then ran the amplified products on a 0.75% agarose gel to confirm successful isolation of the fragments. Next, using a QIAquick Gel Purification Kit (Qiagen), we extracted and purified the 12 fragments directly from agarose to remove any excess primers, nucleotides and buffers. We then sequenced the 5' and 3' ends of all fragments before restriction digestion to re-confirm the positive isolation of the target regions, and to identify any errors that might have been introduced by the polymerase.

Plasmid construction and insert ligation

We cloned all LR-PCR amplified sequences into pGL4.10 basic promoter-reporter vectors (Promega). These vectors contain a synthetic firefly luciferase coding

sequence (*luc2*), but no eukaryotic promoter or enhancer sequences. Thus, the expression levels of the luciferase gene, driven by either *AVPR1A* or *OXTR* promoter variants, can be empirically quantified and compared to one another. We cleaved the *AVPR1A* fragments (at the primer tails) using Xho I (5') and Bgl II (3') restriction enzymes, and then ligated them to pGL4 vectors. We then ligated the *OXTR* fragments into the vectors using Sfi I (5') and Hind III (3') restriction enzymes. Vector maps of the tested constructs are shown in **Figure 6.4**. We used a "no-cassette" vector (pGL4.10 without insert) to act as a negative control for downstream transfection and expression experiments. Vector ligation and transformations were performed at the Sequencing Core Facility in the Department of Genetics at Penn.

Multiple Cloning Region:



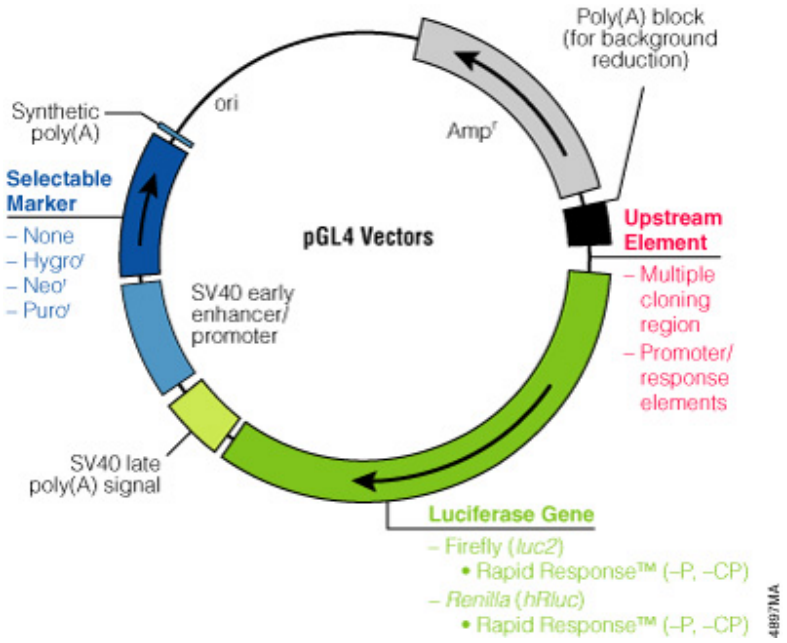


Figure 6.4. Promega pGL4 vector plasmid design, featuring a multiple-restriction enzyme cloning region.

After incubation with restriction enzymes, we treated digested vectors with Antarctic phosphatase to remove 5` phosphate groups, thus preventing any possibility of self-ligation. We then ran both de-phosphorylated vector and digested constructs on a 1% agarose gel. Once again, using a QIAquick Gel Purification Kit (Qiagen), we extracted and purified the digested fragments from agarose. To perform the ligation reaction, we quantified the gel-purified plasmid and the inserts by spectrophotometry (NanoDrop), combined the cleaned double digested constructs and plasmids at a molar

mass ratio of 3:1 (vector:insert) in a reaction using T4 DNA ligase (New England Biolabs) and adenosine triphosphate (ATP). All controls were prepared in the same manner.

Transformation

Using our 12 recombinant constructs (ligated vector+insert), we transformed competent cells (DH5α). DH5α are modified *E. coli* cells that are capable of being efficiently transformed by large plasmids using a heat shock protocol. We made two Luria Bertani (LB) agar plates for each ligation-transformation reaction. In each plate, we added 5μL of ligation reaction to 50 μl of DH5α competent cells, following the appropriate preparation of LB and related reagents (Inoue et al., 1990).

After 24 hours, we inspected all plates for growth, and successful colonies were collected for further testing. We also inspected the vector and insert control plates, as no growth should be observed on these plates. From each plate processed in this manner, we selected ten colonies for extraction and downstream analysis.

From the selected colonies of transformed cells, we extracted ultra-pure plasmid DNA suitable for transfections with a HiSpeed Plasmid Maxi Kit (Qiagen). Plasmid DNA is extracted from *E. coli* by subjecting the cells to alkaline lysis, followed by the binding of plasmid DNA to an ion-exchange column. Once bound, the DNA is purified through the removal of proteins, RNA, and other impurities by using a low salt buffer. Because any growth on vector control plates might indicate contamination, incomplete restriction or re-ligated restricted vectors, we double-digested the extracted plasmid DNA to confirm that the construct recombined with the plasmid. Following incubation with

the BamH I restriction enzyme, we ran the digested products on 1% agarose gel for 20 minutes at 100 V with variable current and visualized the gel to identify successfully transformed colonies. Colonies that were successfully transformed with vector-containing insert will appear to have two bands following digestion. We discarded any clones that did not cut appropriately.

During this phase of the experiment, we experienced several methodological hurdles, particularly with the *OXTR* fragments, that could be directly attributed to the sensitive restriction enzymes (Sfi I) that we used to cleave and ligate the regions to the vectors. The composition of the *OXTR* region was such that, despite three experimental re-designs, very few of the restriction sites present in all six species didn't also cut somewhere else inside the desired fragment. With only 13 standard restriction sites from which to choose in the multiple-cloning region within the pGL4.10 vectors (**Figure 6.4**), it was necessary to utilize the Sfi I enzyme and cut its sites if we were to retain long, truly orthologous comparative regions across species. We found that the Sfi I enzyme was quite sensitive to potential dcm methylation, its activity was blocked by CpG methylation, and it lost efficiency in the presence of basic transformation buffers and conditions.

To verify the orientation of the inserted DNA and amplification fidelity, we directly sequenced all positive clones using internal pGL4 primers (Promega) and BigDye^{TM} v3.1 (ABI) sequencing chemistry. Excess dye terminators were removed with the BigDye XTerminator^{TM} purification kit (ABI), and DNA sequences were read on a 3730xl Gene Analyzer (ABI). In preparation of transfection of our plasmid DNA constructs into mammalian cells, we measured DNA concentration and quality using a

NanoDrop ND-1000 spectrophotometer (Thermo Scientific). All final DNA concentrations were then diluted to fall between 200 – 400 μg/mL with a DNA:protein (A_{260}:A_{280}) ratio of 1.7-1.9.

Cell Line Transfection

For our transfection experiments, we used cells from the human neuroblastoma cell line SH-SY5Y (American Type Culture Collection). The neuroblastoma cell lines act as physiological proxies for the neuronal cells that normally express *AVPR1A* and/or *OXTR* in the brains of mammals, but must be transiently transfected with a bacterial vector to introduce foreign genetic material. The cells were stored and grown according to previously published protocols (Sabol et al., 1998; Hammock and Young, 2004; Tansey et al., 2011) at the Cell Culture Core Facility in the Department of Genetics at Penn. Cell growth was maintained with Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and using collagen-coated flasks.

We counted cells using a haemocytometer to assess their density in a mixed cell suspension. We then seeded cells into 24-well plates at a density between 1000-10,000 cells/well, which we then confirmed using a light microscope with 100X magnification. These plates were maintained in RPMI 1640 and EMEM/F12 medium containing 10% FBS and 2% L-glutamic acid, then incubated at 37°C and 5% CO2.

Twenty-four hours later, we transiently co-transfected each well of plated cells with a firefly luciferase test plasmid (pGL4.10 + insert) and a renilla luciferase plasmid (pGL4.74 with a TK promoter, Promega) at a 2.5:1 (firefly:renilla) ratio. By co-transfecting the cells using renilla luciferase (*hRluc*), an enzyme naturally produced by

the bioluminescent sea pansy (*Renilla reniformis*), we were provided with a secondary luminescent read that could be used to normalize firefly luciferase expression across all experiments. In addition, each 24 well plate contained three replicates of each of an empty pGL4.10 vector (*luc2* with no promoter), a positive control pGL4.13 vector (*luc2* with SV40 promoter), a normalizing control of only pGL4.74 vector (*hRluc* and TK promoter), a negative control consisting only of cells and transfection reagent, and a negative control containing only cells. We performed all transfections in triplicate in four separate batches, using the FuGENE transfection reagent (Promega) according to the manufacturer's instructions. Following the addition of the FuGene transfection reagent at a ratio of 3:1 (reagent:DNA), we placed the cells in the incubator at 37°C with 5% CO_2 for 48 hours.

Dual-luciferase reporter assay

The Dual-Luciferase® Reporter (DLR) assay (Promega) provides an efficient way to assess promoter activity by exploiting the distinct evolutionary origins of firefly and renilla luciferases, and their requirements for dissimilar substrates. In this system, the luminescence of the reporter firefly luciferase is measured first and eventually quenched while simultaneously activating the luminescent reaction of renilla luciferase. We performed all DLR assays according to the manufacturer's instructions.

Prior to reading luminescence, we aspirated all cell medium and transfection reagents from the 24-well plates, and washed each cell monolayer with 500 uL 1X phosphate buffered saline (PBS). Upon the removal of PBS, we used PLB lysis buffer (Promega) to passively lyse the cells directly in the 24-well plates, and placed each plate

on an orbital shaker for 15 minutes. Following cell lysis, we transferred 20 uL of cell lysate from each well into corresponding wells on an opaque sided, clear bottom luminometer plate containing pre-aliquoted 100 uL per well volumes of LAR II luciferase buffer (Promega). We immediately measured luminescence on a GloMax Multi-Detection plate-reading luminometer (Promega). Each well was read for 10 seconds, and the entire plate was read in five replicate sessions with a 1-minute hold in between each replicate.

To quantify the background TK-driven renilla luminescence in each transfection experiment, we quenched the luciferase reactions and promoted renilla luciferase production by injecting 100 uL of Stop & Glo® reagent (Promega) into each well, and re-reading the plate on the luminometer. Once again, each well was read for 10 seconds, and the entire plate was read in five replicate sessions with a 1-minute hold in between each replicate. All readings were recorded automatically, and we corrected for transfection efficiency by normalizing firefly luminescence relative to the luminescence of the co-transfected renilla luciferase and the empty vector controls.

Analysis of luminescence

We measured firefly luciferase luminescence (as relative fluorescent units, RFUs) ten times apiece for each of the eight constructs, and recorded renilla luciferase luminescence values that were generated by the co-transfected pGL4.74 vectors. On each 24-well plate that was prepared, we ran three wells each of two positive controls (pGL4.13, pGL4.10) and two negative controls (cells and FuGENE, cells only). We calculated the mean firefly luminescence values for every transfection reaction of each

construct, and the mean renilla luminescence for each batch of experiments. Next, we calculated a mean instrument background luminescence value (firefly: 241.2 RFUs, renilla: 378.3 RFUs) for the GloMax luminometer by reading empty wells, and subtracted this amount from each of the constructs' mean firefly luminescence value, and from each of the batches' mean renilla luminescence value. From these instrument-corrected means, we expressed the overall luminescence of a given construct as the ratio of its mean firefly luminescence to mean renilla luminescence. Lastly, the ratios were normalized in relation to the luminescence values of the empty vector transfections (pGL4.10 without insert), which was designated as having a ratio of 1.0. The resulting luciferase expression values, as produced by the different *AVPR1A* and *OXTR* promoters of the six primates being examined, were then used in comparative analyses to assess the effect of promoter composition on expression variation.

6.3 Results

AVPR1A promoter sequence diversity

AVPR1A promoter regions are highly variable across primate species (**Table 6.4**). We found high levels of diversity in the *AVPR1A* promoter (5' regulatory region site identity: 66.9% of 4571 bp) compared to the coding region portion of the same fragment (88.5% of 662 bp). Moreover, the 5' regulatory region was characterized by substantial numbers of insertion/deletions (32) in different primate lineages (**Figure 6.2**). Thus, within the entire 5233 bp matrix (gapped), only 3644 sites were identical (69.6% site identity), despite an 84% overall pairwise identity of the full alignment. The bulk of this diversity derived from clade-specific (e.g., platyrrhine-only, hominoid-only) insertions or

deletions of large strings of nucleotides in the region upstream of the methionine start codon (ATG), highlighting the molecular instability of this non-coding region across primate genomes.

Table 6.4. Summary statistics for *AVPR1A* and *OXTR* LR-PCR fragments in 6 primate samples.

	AVPR1A		
	Full fragment 5233 bp	**5' Regulatory 4571 bp**	**Coding 662 bp**
Sequence diversity:			
Pairwise identity of alignment (%)	84.0%	82.3%	94.9%
GC content (% non-gap)	42.8%	39.5%	65.5%
Identical sites	3644 (69.6%)	3058 (66.9%)	586 (88.5%)
Polymorphic sites	1589	1513	76
Insertion/deletions >5 bp	32	32	0

	OXTR		
	Full fragment 5914 bp*	**5' Regulatory 4860 bp**	**Coding 1054 bp***
Sequence diversity:			
Pairwise identity of alignment (%)	88.5%	86.8%	95.9%
GC content (% non-gap)	47.5%	43.7%	65.1%
Identical sites	4352 (73.6%)	3401 (70.0%)	951 (90.2%)
Polymorphic sites	1562	1459	103
Insertion/deletions >5 bp	26	26	0

Note: Each LR-PCR fragment (~5.2-5.9 kb) contains approximately 4.5-4.9 kb of 5' regulatory sequence, and either 662 bp (*AVPR1A*) or 1054 bp (*OXTR*) of coding sequence.
* = The "coding" portion of the *OXTR* fragments for the different primate species contains all of *OXTR* exon 1 (922 bp), plus the 5' end of the intron (132 bp).

All the platyrrhine taxa had a complex GATA/GACA microsatellite repeat sequence motif within the *AVPR1A* 5' regulatory region. This pattern distinguished the platyrrhines from their catarrhine and hominoid relatives, all of which possessed a GATA motif in same RS1 location (Donaldson et al., 2008; Rosso et al., 2008).

OXTR promoter sequence diversity

The promoter region of the *OXTR* also displayed marked sequence variability across primate species (**Table 6.4**). We observed high levels of *OXTR* promoter diversity (5' regulatory region site identity: 70.0% of 4860 bp) compared to the *OXTR* coding region sequences from those same individuals (90.2% of 1054 bp). In the entire 5914 bp matrix (gapped), 4352 sites were identical (73.6% site identity), and there was 88.5% overall pairwise identity of the full alignment. The pairwise identity value for the 5' region immediately upstream of the *OXTR* methionine start codon (ATG) was 86.8%, whereas the coding region had a pairwise identity value of 95.9%. Nevertheless, the overall diversity of the 5' regulatory region of *OXTR* was noticeably less than what we observed across the *AVPR1A* regulatory regions of different primate species.

In addition, we recognized a 540 bp insertion in *Pan troglodytes* (**Figure 6.3**). This chimp-specific insertion, located ~3 kb upstream of the ATG start site, was characterized by a heterogeneous sequence of nucleotides that was slightly A+T rich. Although the 540 bp insertion does not appear to directly code for proteins (as assessed by the protein prediction plug-in in Geneious), it does provide seven additional potential binding sites (of >6 bp) for transcription factors, none of which are present in the other primate *OXTR* regulatory regions.

By contrast, the *OXTR* 5' regulatory sequences of the New World monkeys (*Aotus* and *Callithrix*) were characterized by 10 platyrrhine-specific deletions of 5 bp or greater (from a total of 26 indels >5bp in the entire alignment). One such platyrrhine-specific deletion was 30 bp long, and removed the transcriptional start site (TSS) that putatively functions to initiate transcription in other primate species. We did not observe

this particular 30 bp TSS deletion in any of the non-platyrrhine sequences that we analyzed.

Vector ligation, transformations, and transfections

We were able to generate ~5.2 kb *AVPR1A* promoter amplicons and ~5.9 kb *OXTR* promoter amplicons for each of the six primate species for the purpose of quantifying expression values for 12 experimental plasmid constructs (**Figure 6.4**). However, of these 12, we were only able to ligate 10 fragments to the multiple-cloning region of the pGL4.10 promoter-less vectors (**Table 6.5**). Furthermore, of these 10, we were only able to effectively and reproducibly transform eight of the original constructs. These successful transformations included *Aotus*, *Callitrhix*, *Macaca*, *Pongo* and *Homo* for the *AVPR1A* regulatory region, and *Aotus*, *Macaca* and *Homo* for the *OXTR* regulatory region.

Table 6.5. Vectors and constructs used in transfection experiments.

Project No.	Vector/Construct	Product No.	Success
1	Firefly *luc2* vector + promoter of *OXTR* gene from *Aotus azarai*	pGL4.10	Yes
2	Firefly *luc2* vector + promoter of *OXTR* gene from *Callithrix geoffroyi*	pGL4.10	No
3	Firefly *luc2* vector + promoter of *OXTR* gene from *Macaca mulatta*	pGL4.10	Yes
4	Firefly *luc2* vector + promoter of *OXTR* gene from *Pongo pygmaeus abelii*	pGL4.10	No
5	Firefly *luc2* vector + promoter of *OXTR* gene from *Pan troglodytes*	pGL4.10	No
6	Firefly *luc2* vector + promoter of *OXTR* gene from *Homo sapiens*	pGL4.10	Yes
7	Firefly *luc2* vector + promoter of *AVPR1A* gene from *Aotus azarai*	pGL4.10	Yes
8	Firefly *luc2* vector + promoter of *AVPR1A* gene from *Callithrix geoffroyi*	pGL4.10	Yes
9	Firefly *luc2* vector + promoter of *AVPR1A* gene from *Macaca mulatta*	pGL4.10	Yes
10	Firefly *luc2* vector + promoter of *AVPR1A* gene from *Pongo pygmaeus abelii*	pGL4.10	Yes
11	Firefly *luc2* vector + promoter of *AVPR1A* gene from *Pan troglodytes*	pGL4.10	No
12	Firefly *luc2* vector + promoter of *AVPR1A* gene from *Homo sapiens*	pGL4.10	Yes
Control	Firefly *luc2* vector without promoter insert ("empty cassette")	pGL4.10	n/a
Control	Firefly *luc2* vector + SV40 promoter	pGL4.13	n/a
Control	Renilla *hRluc* vector + TK promoter	pGL4.74	n/a

Despite these limitations, we were able to produce robust transformations at high yields for each of the eight fragments that successfully ligated to the pGL4.10 vectors. We positively confirmed the sequence integrity of the inserts through restriction enzyme (using BamHI) digestion followed by gel electrophoresis and cut-site analysis. All eight constructs were successfully transfected into the SH-SY5Y neuroblastoma cells in four independent replicate batches.

Luciferase assay and expression quantification

For the *AVPR1A* promoter-reporter assays, we observed differential activity for the promoters of the five primate species when expressed as a luciferase constructs in pGL4 control vectors. In SH-SY5Y cells, the *AVPR1A* promoter from *Pongo pygmaeus abelii* showed 18 times higher relative activity compared to the empty vector (**Figure 6.5**). In fact, all of the *AVPR1A* promoter constructs, with the exception of *Callithrix geoffroyi*, exhibited markedly greater activity than the controls.

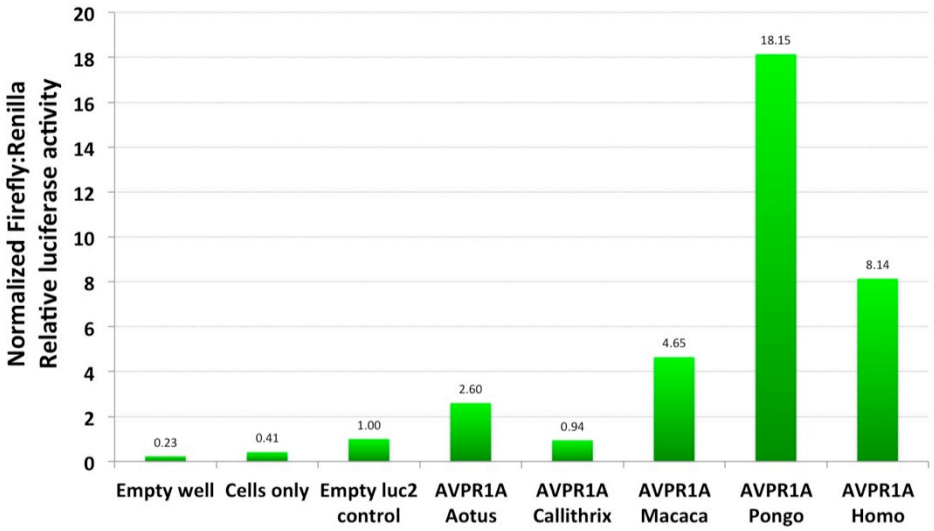


Figure 6.5. *AVPR1A* luciferase ratio luminescence histogram chart depicting expression values of different primate (and control) *AVPR1A* promoters.

For the *OXTR* promoter-reporter assays, we also observed clearly differential activity across the promoters of three primate species when expressed as a luciferase constructs in pGL4 control vectors. When transfected into SH-SY5Y cells, the *OXTR* promoter from *Aotus azarai* showed 1.22 times higher relative activity compared to the empty pGL4.10 vector (**Figure 6.6**). In addition, the *Macaca mulatta* exhibited 1.75 times more activity than the control, and the *Homo sapiens* construct 1.88 times more luciferase expression.

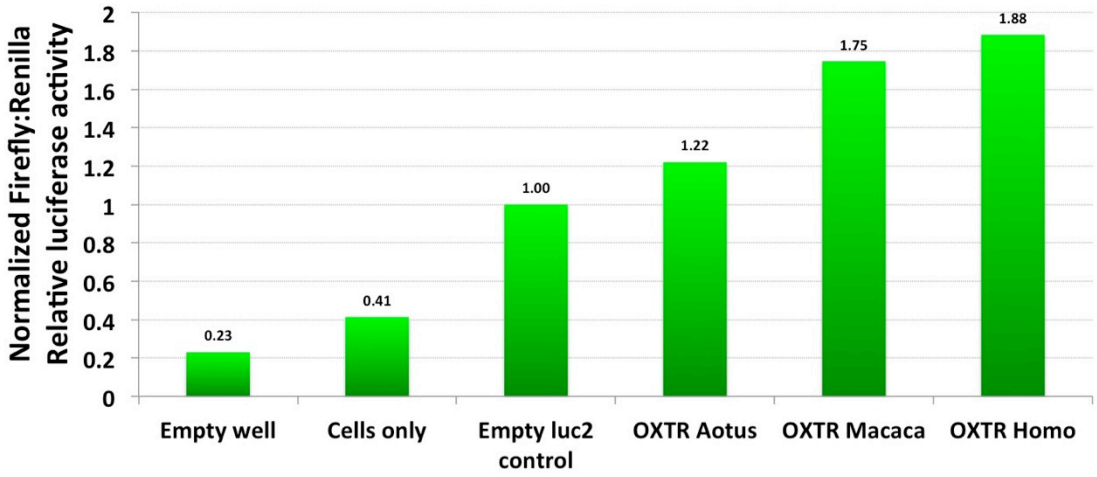


Figure 6.6. *OXTR* luciferase ratio luminescence histogram chart depicting expression values of different primate (and control) *OXTR* promoters.

6.4 Discussion

AVPR1A expression variation in primates

When inserted into pGL4.10 vectors and transfected into SH-SY5Y cells, the regulatory regions of *AVPR1A* from five different primate taxa expressed significantly different amounts of *luc2*. Although each species has accumulated notable amounts of molecular variation in its respective *AVPR1A* regulatory region, the fragments that we examined were directly orthologous to one another, and occupied the same 5.2 kb genomic "footprint" upstream of the *AVPR1A* gene in each taxon. Thus, it is reasonable

to assume that *AVPR1A* expression varies widely across primate lineages. This variation can be directly traced back to the sequence composition (insertions, deletions, repeats, and single nucleotide polymorphisms [SNPs]) of the *AVPR1A* regulatory region belonging to different taxa.

However, it should be noted that we were not able to formally test all of the molecular regions relevant to *AVPR1A* expression in primates. In particular, RS3 and GT_{25}, present in Old World monkeys and apes (and absent in strepsirhines and New World monkeys), lie just outside of the fragment that we isolated and assessed in all species (see Chapter 4). This region was excluded from our study in order to retain orthology across the sequences being compared (platyrrhines do not posses these repeat regions). Therefore, our conclusions on expression differences are based on the species-specific SNPs, insertion/deletion events, and RS1 sequences and repeats. In addition, intraspecific variation at the *AVPR1A* regulatory region was not assessed by this study, meaning that higher (and lower) *AVPR1A* expression levels are likely to exist within the species we compared in this broad phylogenetic analysis.

Given our findings, and considering the caveats, it appears that RS1-based *AVPR1A* expression does not directly correlate with pair bonding in primates, as the promoters belonging to solitary orangutans appear to express *luc2* at the highest levels (18x control). Further studies must necessarily include many more species, and assay a complete range of all intra- and interspecific repeat sequences, in order to fully evaluate the function of *AVPR1A* on primate social behaviors.

OXTR expression variation in primates

When similarly inserted into pGL4.10 vectors and transfected into SH-SY5Y cells, the regulatory regions of *OXTR* from three different primate taxa express *luc2* at much less variable rates. Like *AVPR1A*, each species has accumulated notable amounts of molecular variation in its respective *OXTR* regulatory region (see Chapter 5), and the 5.9 kb fragments that we examined were directly orthologous for each of the different primate species. However, the variation in expression of the vector constructs bearing the *OXTR* promoters of owl monkeys, macaques, and humans did not occur at the same magnitude observed with the *AVPR1A* constructs.

Nevertheless, the *OXTR*-related *luc2* expression was quantifiably different across species, highlighting a direct relationship between the sequence composition of regulatory regions and heightened *OXTR* expression. Humans appear to express *luc2* at higher rates, pointing to the enhancement of promoter activity through the human-specific sequence composition upstream of the *OXTR* gene. However, like the case of the *AVPR1A* promoter, there exists another repeat region further upstream of the region we were able to isolate and examine (see Chapter 5). In most species, this region consists of two (*Aotus*, *Callithrix*) to five (*Pongo*, *Gorilla*, *Nomascus*, *Macaca*) dinucleotide repeats the sequence A+T. In chimpanzees, this region consists of roughly 32 A+T repeats, and in humans there are over 120 A+T repeats. Such sequence variation might result in even greater enhancement of *OXTR* expression in humans and chimps, and may explain the reduced levels of expression that we observed in this study. Finally, it would be greatly desirable to compare *OXTR* promoter functionality in more than just the three primate species for which we were able to collection expression data.

Conclusions

The results from both *AVPR1A* and *OXTR* luciferase gene reporter assays are intriguing, with different sequence motifs leading to significant variation in *AVPR1A* or *OXTR* mRNA production in different primate species. Such species-level sequence variation might help explain the diversity and intensity of pro-social behaviors in different primate taxa. However, our results do not point to a simple pattern of expression for *AVPR1A* and *OXTR* that is strongly associated with partner preference in different primate species. Monogamous owl monkeys, for example, show lower levels of expression at both genetic loci. Thus, other mechanisms must be involved in the manifestation of primate pair bonds.

In addition, in the OXTR promoter there is a platyrrhine-specific deletion 30 bp long that removes the TSS, which purportedly functions to initiate transcription of the gene for other taxa. Such a molecular event strongly suggests that the platyrrhine TSS has relocated elsewhere, so that *OXTR* transcription can still be initiated in platyrrhine species. Despite these changes, there appears to be greater synteny across the primate chromosomes at the *OXTR* locus, and genetic material does not appear to have crossed-over from elsewhere in the genome, as may have been the case at the *AVPR1A* promoter.

The OT and AVP pathways are extremely complex, and intersect with other neurological pathways, such as the dopaminergic or serotinergic pathways, meaning that even slight modulations at any point along these systems could influence downstream behavioral phenotypes. Therefore, there must be many, many different points along the genomes of primates that have undergone functional changes, and ultimately influenced the differences in primate pair bonding behaviors.

Chapter 7: Comparisons and Conclusions

7.1 Restatement of the problem

Primates are social organisms. Throughout their evolution, primates have adopted a wide range of social behaviors, and are organized socially across a full spectrum of mating systems. Of these social and mating systems, 'monogamy' is exceedingly rare in primates (van Schaik and van Hoof, 1983; van Schaik and Dunbar, 1990). In addition, monogamous social systems are widely distributed across the primate order, indicating that they have arisen independently multiple times (Di Fiore and Rendall, 1994; Rendall and Di Fiore, 2007).

Primate social behaviors find their biological basis in neuronal cell-signaling pathways and the neuroendocrine system. Although influenced by external stimuli, these systems are mediated by endogenous chemicals (and their receptors) that are blueprinted by genes. Thus, molecular changes that alter the structure or expression of such genes can functionally affect the behavioral phenotypes.

This scenario raises many important questions. *Are the social behaviors of primates regulated by the same biological pathways in different species? What genetic changes were necessary to create variable sociobehavioral biological pathways from pre-existing genetic machinery? How did these molecular evolutionary events take place, and when did they occur? Are these changes the result of drift or other population genetic events? What can these changes say about the directionality of selection?*

The evolution of social behaviors must have its origins in the molecular changes to the structure and expression of the neurogenetic loci, which then significantly affected their function (Enard & Pääbo, 2004). Most behavioral phenotypes are polygenic in

origin, and thus represent a large proportion of the genome at which mutational or recombination events might accumulate. SNPs and other small mutations can alter the product of genes through novel, gain-of-function mutations, while mutations within the regulatory regions or the presence of large genomic CNVs can likewise create new expression patterns (Perry et al., 2007; Sebat et al., 2007; Lee et al., 2008). When viewed together, these molecular changes could certainly have provided the variation required for both slow- and fast-paced evolution and selection of phenotypic changes in social behavior.

By comparing genetic regions that are related to these behaviors across primate taxa, one can begin to determine whether the evolution of complex social behavior occurs through conserved molecular mechanisms. Similarly, the directionality of selection can be inferred from patterns of sequence divergence in candidate genetic loci, and used to generate explanations for the evolution of these loci and their related behaviors. Data generated in this fashion may also provide the necessary context to identify the selective forces (ecological, social, pathological) involved in the development and maintenance of complex social behaviors like mate selection.

This dissertation has explored questions regarding the evolution of pro-social behaviors through the surveying of neurogenetic variation in *AVPR1A* and *OXTR* in a wild population of monogamous Azara's owl monkeys (*Aotus azarai azarai*), and other platyrrhines. Azara's owl monkeys exhibit social behaviors such as the formation of pair bonds and significant paternal care to offspring that are exceedingly rare among the majority of primate taxa. Therefore, owl monkeys may serve as a useful model for investigating the relationship between molecular variation at these neurogeneic loci and

the social behaviors (and mating systems) exhibited by different lineages of primates. Thus, the goal of this study has been to examine the molecular evolution of *AVPR1A* and *OXTR* in owl monkeys to better understand how the pro-social behaviors related to those loci may have evolved.

7.2 Owl monkey phylogeographic and evolutionary histories

Arising from one or several small colonization events roughly 35 million years ago, the New World primates (members of the infraorder *Platyrrhini*) have developed a suite of behavioral patterns that differentiate them from their Old World primate relatives (Chapter 1; Poux et al., 2006). Owl monkeys (*Aotus* spp.) exhibit many of these behavioral patterns, including higher degrees of paternal care and monogamous sociality. But how did this distinction between Old World and New World primates arise? How could different patterns of sociobehavioral tendencies develop in owl monkeys and other monogamous platyrrhines?

Because all platyrrhine lineages coalesce to a very few hypothetical ancestral genomes, it may be possible to identify neurogenetic differences that relate to the evolution of lineage-specific social behaviors. However, in order to properly contextualize functional neurogenetic variation related to such sociobehavioral pattterns, it is necessary to first establish the range of molecular variation occurring at non-related genetic loci (Kimura, 1981). To address this issue, we sequenced the entire mitochondrial genome of two species of *Aotus* (*A. azarai* and *A. nancymaae*), and analyzed 39 haplotypes of the mitochondrial COII gene from ten different owl monkey taxa (Chapter 2).

Our analyses point to an old and complex evolution of the genus *Aotus*. Based on our analyses of mtDNA variation, we estimate that TMRCA for all extant owl monkey species is 8.95 mya, a date that is considerably older than the previous coalescent date estimates of 3.6-4.7 mya (Ashley and Vaughn, 1995; Plautz et al., 2009). Yet, an older time frame for the diversification of the genus would also fit the 11.8-13.5 mya estimates for the *Aotus dindensis* fossils, which exhibit traits ancestral to all modern *Aotus* taxa.

At the same time, we also observe a deep phylogenetic split between *Aotus* species living north of the Amazon River from those living to the south, which began over 8 mya. Our data further suggest that *A. nancymaae* was the first owl monkey taxon to diverge from the northern clade around 6.38 mya, once the north-south split had occurred. These clades also correlate with differences in malarial resistance and pelage coat color observed in species from those geographic regions (Ford, 1994; Defler and Bueno, 2007; Fernandez-Duque, 2011).

Furthermore, our findings suggest that the southern expansion of *Aotus* was gradual (e.g., only one chromosomal fission event and the maintenance of the Y-autosomal fusion event in southern males), with species diversifying steadily in different points in time, not through multiple splits or population bottlenecks (Pieczarka et al., 1993, 1998; Torres et al., 1998). This scenario would fit with the paleogeographic history of the South American continent and the Amazon River, including the formation of the gigantic inland Lago Amazonas by the Andean uplift ~9 mya and its drainage ~5.0-2.5 mya, as well as the more recent establishment of southern rivers and the draining of the South American Chaco (Rosenberger et al., 2009).

In support of this interpretation, the majority of southern species are

karyotypically identical. All of them possess 49(male)/50(female) karyotypes, with the exception of *A. nigriceps* (51m/52f). Species found north of the Amazon possess a wide range of karyotypes ranging from 46 in *A. vociferans* to 56 in *A. lemurinus* (Defler and Bueno, 2007). The greater antiquity of northern *Aotus* taxa would have allowed for the many karyotypic changes seen in different owl monkey species.

In conclusion, our data strongly indicate that owl monkeys have a very ancient and complex evolutionary past that has likely been driven by a range of ecological pressures and historical events.

7.3 Owl monkey population history

Investigations into the functional evolution of genetic loci related to social behavior in a given species or population directly benefit from prior knowledge of molecular variation and genetic structure of that taxa (Di Fiore, 2003). Therefore, to understand the recent evolutionary history and genetic structure of our focal owl monkey population, we assessed variation of the mtDNA control region (D-loop) in 118 wild individuals (Chapter 2). Overall, we detected a high level of mtDNA diversity within our wild population of *A. a. azarai*. The majority of individuals (115/118) belong to one of three major CR haplogroups or clades.

Hg-A is the most diverse of the three. It is present in most owl monkey social groups, and based on ρ coalescence estimates for each of the three haplogroups, hg-A is likely to be the ancestral lineage for this population. This finding is further reinforced by the phylogenetic affinity of hg-A to the most distantly related haplotypes X, Y, and Z.

The patterning of the *A. a. azarai* mitochondrial networks and systematics, along with the general agreement among different summary statistics (*D*, F_S, τ, and π), reinforces the possibility that this population has undergone several distinct expansion events in its history, not just a single recent expansion. The relative ages of the mitochondrial clades are consistent with the climatic and geographic processes that drained the southernmost Chacoan forests and flatlands of water some 5,000-7,000 years ago (Iriondo, 1984, 1993). Thus, our haplogroup age estimations indicate that the expansion of the population into the Pilagá watershed soon after that period. This scenario suggests that both putative *A. azarai* subspecies, *A. a. boliviensis* and *A. a. azarai*, had common origins further north, and only recently moved southward, in multiple waves, into the South American Gran Chaco.

Furthermore, to establish our knowledge of genetic kinship and individual identity within the wild population, we investigated autosomal variation in the form of 24 STR microsatellite loci (Chapter 3). Of these loci, 9 were monomorphic and 15 were polymorphic in the 129 individuals that were genotyped. When combined, the 15 polymorphic loci provided a parentage exclusionary percentage of 85.2% when both parents were unknown, and 98.4% when one parent was unknown. These percentages increased to 99.9% when both parents were known, to 100% for the detection of individual identity, and to 99.9% for the estimation of sibling identities.

Our observations from the field have shown that, although male owl monkeys invest significant resources in parenting (Rotundo et al., 2005; Wolovich et al. 2008), it is possible that these resources may be sometimes directed at offspring whom they did not sire (Fernandez-Duque et al., 2008; Fernandez-Duque et al., 2009). Such a finding would

represent a clear deviation from the predictions of kin selection and parental investment theories (Hamilton, 1964a, b; Trivers, 1972). However, our preliminary (and ongoing) analysis of kinship and parentage of wild owl monkeys using this STR panel has revealed that not a single case of extra-pair paternity has occurred in any of the social groups (n=25) under regular observation. Therefore, in owl monkeys, it seems that social monogamy occurs together with genetic monogamy. This finding substantiates the pro-social behavioral phenotype of owl monkeys, and reinforces their value as a case study of the molecular evolution of genetic loci, such as *AVPR1A* and *OXTR*, that are believed to functionally influence traits involved in pair-bonding behaviors and monogamous social systems in primates.

7.4 *AVPR1A* variation in owl monkeys and other primates

Coding region evolution: changes to AVPR1A protein structure

A. azarai has a unique genic structure for *AVPR1A* that varies in coding sequence relative to other primate and mammalian species. Namely, the *A. azarai* sequence exhibits five amino acid changes that may putatively increase the surface area of the ligand-binding domain.

Platyrrhine species have accumulated many synonymous and non-synonymous mutations in the coding sequence of *AVPR1A* when compared to Old World monkeys and hominoid primates (Chapter 4). Each of the six platyrrhine genera examined in this study (*Aotus*, *Lagothrix, Saguinus*, *Callithrix*, *Pithecia* and *Saimiri*) exhibits a number of lineage-specific mutations that distinguishes its *AVPR1A* coding sequences from those of any other taxon being investigated. In addition, all platyrrhines share 24 nucleotide

changes (of the 563 nucleotides in our interspecies comparison) that separate their clade from that of other primates.

Our analyses of primate *AVPR1A* coding sequences suggest that non-neutral evolution, in the form of selection, may be at work at this locus. The high positive value of Tajima's *D* and negative value of Fu's F_s reveal an excess of intermediate frequency alleles among the collection of primate taxa. This observation implies that the observed variation among primate *AVPR1A* sequences is higher than what would be expected based on mutation rate alone, and therefore has likely been shaped by balancing selection (Tajima, 1989a, b; Fu, 1997).

To better understand the distribution of these alleles considering the mixed nature of our interspecific data, we chose to distribute the *AVPR1A* variation on our established phylogenetic topology. In doing so, we observed a slightly different pattern of variation. When considering the ratio of non-synonymous (d_N) to synonymous (d_S) amino acid changes in *AVPR1A*, it appeared that purifying (negative) selection constrained *AVPR1A* variation along the terminal branches leading to many of the different primate lineages. The notable exception to this pattern occurred on the branch leading to the human-chimp-bonobo clade representing the tribe Hominini (d_N/d_S = 12.37). Essentially, the split that separates this hominin clade from all other primates was characterized by many non-synonymous changes, suggesting that diversifying (positive) selection has played a role in the creation of different *AVPR1A* protein isoforms during the early evolution of hominins (see **Figure 4.5**; Chapter 4). Thus, despite high levels of intraspecific coding conservation in 25 *A. azarai* individuals, the *AVPR1A* gene exhibits considerable nucleotide and codon variation among closely related primate species.

Regulatory region evolution: changes to AVPR1A expression

The *AVPR1A* 5' regulatory region has been shown to be highly variable, even in closely related taxa (Kim et al., 2002; Hammock and Young, 2005; Donaldson et al., 2008; Rosso et al., 2008). Our data for the 5' UTR of *AVPR1A* microsatellite variation in different primate taxa clearly support these findings. This putative promoter region has been empirically linked to the functional expression of the *AVPR1A* gene, as well as the density and distribution of V1a receptors in the brain, yet it is tremendously variable across species, and susceptible to major rearrangements and insertion-deletion events. Therefore, it is possible that, in primates (and other mammals), the STR-bearing promoter could be under relaxed selection such that generational expression changes in response to changing social or ecological conditions could take place. Alternatively, the molecular structure of the region could simply be conducive to accelerated mutability in the form of SNPs, in-del substitutions, STR length variants, non-allelic homologous recombination, and even CNV formation.

In our analysis of *AVPR1A* expression (Chapter 6), we noted significant differences in firefly luciferase (*luc2*) transcription in transiently transfected promoter-reporter vectors that were based on the 5' regulatory regions of the gene from different primate taxa. Although many caveats exist regarding the use of these vectors, and the correct isolation of orthologous regulatory elements, we can confidently state that expression is functionally modulated based on the sequence composition of the upstream promoter of *AVPR1A*. Primates exhibit many sequence differences in this region, so it is likely that primates transcribe *AVPR1A* proteins in variable amounts.

Chromosomal evolution: changes in gene dosage, transcriptional regulation, splicing

Our genomic comparison of structural variation at the chromosomal level revealed the conservation of the *AVPR1A* gene in all species compared, although, in many species, major chromosomal rearrangements have apparently reshuffled the regulatory regions upstream of the gene. From the DGV, we noted that, for many primate taxa, CNVs exist in large sections of the entire region (Iafrete et al., 2004). Recent genomic studies have noted that simple variation (STRs), similar to the one present adjacent to *AVPR1A*, may precipitate more mutations, both substitutional and structural (Conrad et al., 2010).

7.5 *OXTR* variation in owl monkeys and other primates

Coding region evolution: changes to OXTR protein structure

A. azarai have a unique genic structure for *OXTR* that varies in coding sequence relative to other primate and mammalian species. Similar to the case of *AVPR1A*, the *A. azarai* sequence exhibits four amino acid changes in the 5' end of the *OXTR* coding region, which may putatively increase the surface area of the ligand-binding domain at the C-terminus of the mature *OXTR* protein. Such changes could qualitatively alter the final structure of the receptor protein known to transmit neuronal social signals (Carter, 1998; Carter et al., 2008).

As we observed with the *AVPR1A* locus, platyrrhine species have accumulated many synonymous and non-synonymous mutations in the coding sequence of *OXTR* when compared to Old World monkeys and hominoid primates (Chapter 5). Each of the seven platyrrhine genera examined in this study (*Aotus*, *Lagothrix, Saguinus*, *Callithrix*,

Callicebus, *Pithecia* and *Saimiri*) exhibits a number of lineage-specific mutations that distinguishes their *OXTR* coding sequences from those of any other taxon being investigated. Moreover, *Callithrix*, *Saimiri* and *Saguinus* possess nearly twice as many amino acid changes as other platyrrhines, and feature a faster rate of mutational accumulation.

Amino acid changes at the *OXTR* locus have also accumulated in different primate species in a lineage-specific manner. The distribution of amino acid changes is notably diverse within the platyrrhine infraorder compared to the localization of non-synonymous changes observed in other taxa. In addition, platyrrhines exhibit the shared derived possession of six amino acid changes in their *OXTR* mRNA that separate their clade from that of other primates. These changes on the receptor could have been a driver of, or a response to, the novel form of the OT neuropeptide that was recently discovered to be present in New World monkeys (Lee et al., 2011).

As for *AVPR1A*, non-neutral selection may be acting on the *OXTR* locus across the Order Primates. The high positive value of Tajima's *D* and negative value of Fu's F_s values reveal an excess of intermediate frequency alleles, which, in turn, suggests that the observed variation is not neutral, and has likely been shaped by balancing selection (Tajima, 1989a, b; Fu, 1997).

Yet, when assessing this variation within the primate phylogenetic tree, we observe a slightly more nuanced pattern of variation. It appears that purifying (negative) selection is constraining *OXTR* variation along the terminal branches leading to each of the different primate lineages. This interpretation is based on the observation that silent nucleotide substitutions (leading to synonymous amino acids, d_S) vastly outnumber

mutations that lead to non-synonymous amino acid changes (d_N) for all primate taxa. By contrast, the split between rodents and primates is characterized by many non-synonymous changes (including one codon deletion), suggesting that diversifying (positive) selection has played a role in the creation of many primate *OXTR* protein isoforms, as evidenced by a d_N/d_S ratio of 7.07 on the branch leading to the entire primate clade (**Figure 5.6**, Chapter 5).

Thus, despite high levels of intraspecific coding conservation in 25 *A. azarai* individuals, the *OXTR* gene exhibits considerable nucleotide and codon between among closely related primate species. This is particularly true for the other platyrrhine primates, as well.

Regulatory region evolution: changes to OXTR expression

Regulatory regions of *OXTR* vary among primates as well, with the insertion of large STR expansions in the *OXTR* 5' UTR/promoters of *Pan troglodytes* (2.5 kb and 4.5 kb upstream of ATG) and *Homo sapiens* (4.5 kb upstream of ATG) being observed. This pattern is in distinct contrast with the 5' UTRs of platyrrhines such as *Aotus* and *Callithrix*, whose promoters are peppered with over 10 deletions of 5 nucleotides (or larger), including a 30 bp deletion that eliminates the putative translational start site of *OXTR*, which is present in all other primate genomes. These Hominini-specific insertions and Platyrhinni-specific deletions could greatly affect the transcriptional efficiency of the *OXTR* gene in those taxa. In fact, in our limited analysis of *OXTR* expression (Chapter 6), we noted significant differences in firefly luciferase (*luc2*) transcription in transiently transfected promoter-reporter vectors that were based on the 5' regulatory regions of the

gene from different primate taxa. Despite ours being a limited comparison between *Aotus-Macaca-Homo*, we still observed distinct differences in the amount of *luc2* that was expressed by the different vector constructs featuring the different promoters of those taxa.

It is therefore reasonable to surmise that such alterations of expression would likely reduce or increase *OXTR* mRNA translation, and subsequently influence the distribution and density of *OXTR* proteins in the brain. Such changes could, in turn, have dramatic affects on signaling capabilities of the OT pathway, and ultimately impact behavior.

Chromosomal evolution: changes in gene dosage, transcriptional regulation, splicing

Our genomic comparison of the *OXTR* locus at the chromosomal level revealed that, despite the presence of SNPs and insertion-deletion events, major chromosomal rearrangements have not reshuffled the regulatory regions upstream of the gene in primates. This observation greatly distinguishes *OXTR*'s evolutionary history from that of the *AVPR1A* locus (Chapter 4), which has experienced a chromosomal rearrangement in the 5' upstream region in platyrrhines. Nevertheless, for many primate taxa, CNVs and other structural changes (segmental duplications, STR length polymorphisms) exist in large sections of the entire region (Gregory et al., 2009).

7.6 Comparisons

Despite the polygenic characteristics of the AVP and OT pathways and high levels of intraspecific sequence conservation of the *AVPR1A* and *OXTR* genes among

Azara's owl monkeys, our findings suggest that both pathways' receptor proteins have undergone a considerable amount of evolutionary change across primate taxa. To directly assess the strength of these observations and interpretations, we performed targeted intra- and interspecific comparisons of both genes to the COII coding region of the mitochondrial genome across 26 primate species and subspecies. This analysis revealed that the coding regions of *AVPR1A* and *OXTR* have evolved in a non-neutral fashion.

We first calculated the variability of nucleotide sites at all three loci in 25 owl monkey individuals (**Table 7.1**). As described previously, there was little intraspecific variation at the COII, *AVPR1A* and *OXTR* loci. In fact, only one SNP was present at the *OXTR* locus in this particular sampling of the larger population (**Table 7.2**). Although it is not surprising that coding genes are conserved within a species, it was reassuring that all three unlinked loci exhibit nearly 100% pairwise identity measures in 25 individuals.

Table 7.1. Intraspecific samples investigated at the mtDNA COII, *AVPR1A*, and *OXTR* genes.

ID	Species	Common Name	Sex	Locale
AaaPLunk[a]	*Aotus azarai*	Azara's Owl Monkey	Unk	Core area, Formosa, AR
Aaa008	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa014	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa015	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa021	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa032	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa034	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa037	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa053	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa057	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa063	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa067	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa071	*Aotus azarai*	Azara's Owl Monkey	M	Downstream, Formosa, AR
Aaa082	*Aotus azarai*	Azara's Owl Monkey	M	Downstream, Formosa, AR
Aaa087	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa092	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa108	*Aotus azarai*	Azara's Owl Monkey	F	Upstream, Formosa, AR
Aaa109	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
Aaa114	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa122	*Aotus azarai*	Azara's Owl Monkey	F	Core area, Formosa, AR
Aaa123	*Aotus azarai*	Azara's Owl Monkey	M	Core area, Formosa, AR
AaaF1	*Aotus azarai*	Azara's Owl Monkey	F	Saenz-Pena Zoo, AR
AaaF1B	*Aotus azarai*	Azara's Owl Monkey	F	Saenz-Pena Zoo, AR
AaaF2	*Aotus azarai*	Azara's Owl Monkey	F	Saenz-Pena Zoo, AR
AaaM2	*Aotus azarai*	Azara's Owl Monkey	M	Saenz-Pena Zoo, AR

[a] = Control *A. azarai* sample used in this study.

Table 7.2. Summary statistics for mtDNA COII, *AVPR1A* and *OXTR* sequences in 25 owl monkey individuals.

	COII mtDNA 696 bp	*AVPR1A* Exon 1 563 bp	*OXTR* Exon 1 922 bp
Sequence diversity:			
Pairwise identity of alignment (%)	100%	100%	99.95%
GC content (% non-gap)	38.8%	65.4%	64.80%
Summary statistics:			
Usable loci (<5% missing data)	696	563	922
Identical sites	696 (100%)	563 (100%)	921 (99.9%)
Polymorphic sites	0	0	1
Single nucleotide polymorphisms (SNPs) [freq.]	-	-	819T>A (34%)

Next, we compared the sequence composition of three genes in the 26 primate taxa for which we had equivalent data (**Table 7.3**). Using only one representative sequence per taxa, we calculated the interspecific nucleotide variation for each gene (**Table 7.4**), and constructed maximum likelihood phylogenetic trees to visually represent this variation at the different loci (**Figure 7.1**). Interestingly, we observed that, across primates, the COII gene appears to be more variable than either of the other two loci. This is true in terms of lower pairwise identity values (77.8%) and the long branches present at every branch of the COII phylogenetic tree.

Table 7.3. Interspecific samples investigated at the mtDNA COII, *AVPR1A*, and *OXTR* genes.

ID	Species	Common Name	Sex	Locale / Source
AaaPLunk[a]	*Aotus azarai*	Azara's Owl Monkey	Unk	Core area, Formosa, AR
Al001	*Aotus lemurinus*	Grey-bellied Owl Monkey	Unk	San Diego Zoo/CRES
Alg002	*Aotus lemurinus grisiemembra*	Grey-handed Owl Monkey	Unk	San Diego Zoo/CRES
Ana001	*Aotus nancymaae*	Nancy Ma's Owl Monkey	Unk	San Diego Zoo/CRES
Av002	*Aotus vociferans*	Spix's or Noisy Owl Monkey	Unk	DuMond Conservancy
Cd001	*Callicebus donacophilus*	White-eared Titi Monkey	M	San Diego Zoo/CRES
Ppp001	*Pithecia pithecia*	White-faced Saki Monkey	M	San Diego Zoo/CRES
Sss001	*Saimiri sciureus*	Common Squirrel Monkey	M	San Diego Zoo/CRES
Hs001 or hg19	*Homo sapiens*	Human	M	Self (P.Babb) / UCSC
Hle001	*Hylobates leucogenys*	White-cheeked Gibbon	M	Coriell Institute (IPBIR: PR01038)
Hsy001	*Hylobates syndactylus*	Siamang	F	Coriell Institute (IPBIR: PR00969)

ID	Species	Common Name	Sex	Locale / Source
Ppa001	*Pan paniscus*	Bonobo (Pygmy Chimpanzee)	F	Coriell Institute (IPBIR: PR00092)
Ptr002 or panTro3	*Pan troglodytes*	Chimpanzee	F	Coriell Institute (IPBIR: PR00605) / UCSC
Pab001 or ponAbe2	*Pongo pygmaeus abelii*	Sumatran Orangutan	M	Coriell Institute (IPBIR: PR00253) / UCSC
Cge001	*Callitrhix geoffroyi*	White-fronted Marmoset	F	Coriell Institute (IPBIR: PR01094)
Cmo001	*Callicebus moloch*	Dusky Titi Monkey	M	Coriell Institute (NIA: NG06115)
Lla001	*Lagothrix lagotricha*	Woolly Monkey	F	Coriell Institute (NIA: NG05356)
Mamu001 or rheMac2	*Macaca mulatta*	Rhesus Macaque	F	Coriell Institute (NIA: NG06249) / UCSC
Sfu001	*Saguinus fuscicollis*	White-lipped Tamarin	F	Coriell Institute (NIA: NG05313)
hg19	*Homo sapiens*	Human	Unk	UCSC Genome Browser
gorGor3	*Gorilla gorilla gorilla*	Gorilla	Unk	UCSC Genome Browser
nomLeu1	*Nomascus leucogenys*	Gibbon	Unk	UCSC Genome Browser
calJac3	*Callithrix jacchus*	Marmoset	Unk	UCSC Genome Browser
mm9	*Mus musculus*	Mouse	Unk	UCSC Genome Browser
rn4	*Rattus norvegicus*	Rat	Unk	UCSC Genome Browser
otoGar3	*Otolemur garnettii*	Bushbaby	Unk	Ensembl

Note: Core study area is located in Formosa Province, Argentina (Lat. = 25º, 59.4 min South; Long. = 58º, 11.0 min West).
[a] = Control *A. azarai* sample used in this study.

Table 7.4. Summary statistics for mtDNA COII, *AVPR1A* and *OXTR* sequences in 26 primate and rodent species.

	COII mtDNA 549 bp	*AVPR1A* Exon 1 563 bp	*OXTR* Exon 1 922 bp
Sequence diversity:			
Pairwise identity of alignment (%)	77.8%	92.2%	94.2%
GC content (% non-gap)	42.3%	66.0%	65.2%
Summary statistics:			
Usable loci (<5% missing data)	549	563	922
Identical sites	254 (46.3%)	362 (64.3%)	675 (73.2%)
Polymorphic sites	295	201	247

AVPR1A: the bushbaby (otoGar1) individual has a 24 bp deletion.
OXTR: both the mouse (mm9) and rat (rn4) individuals have 3 bp deletions.

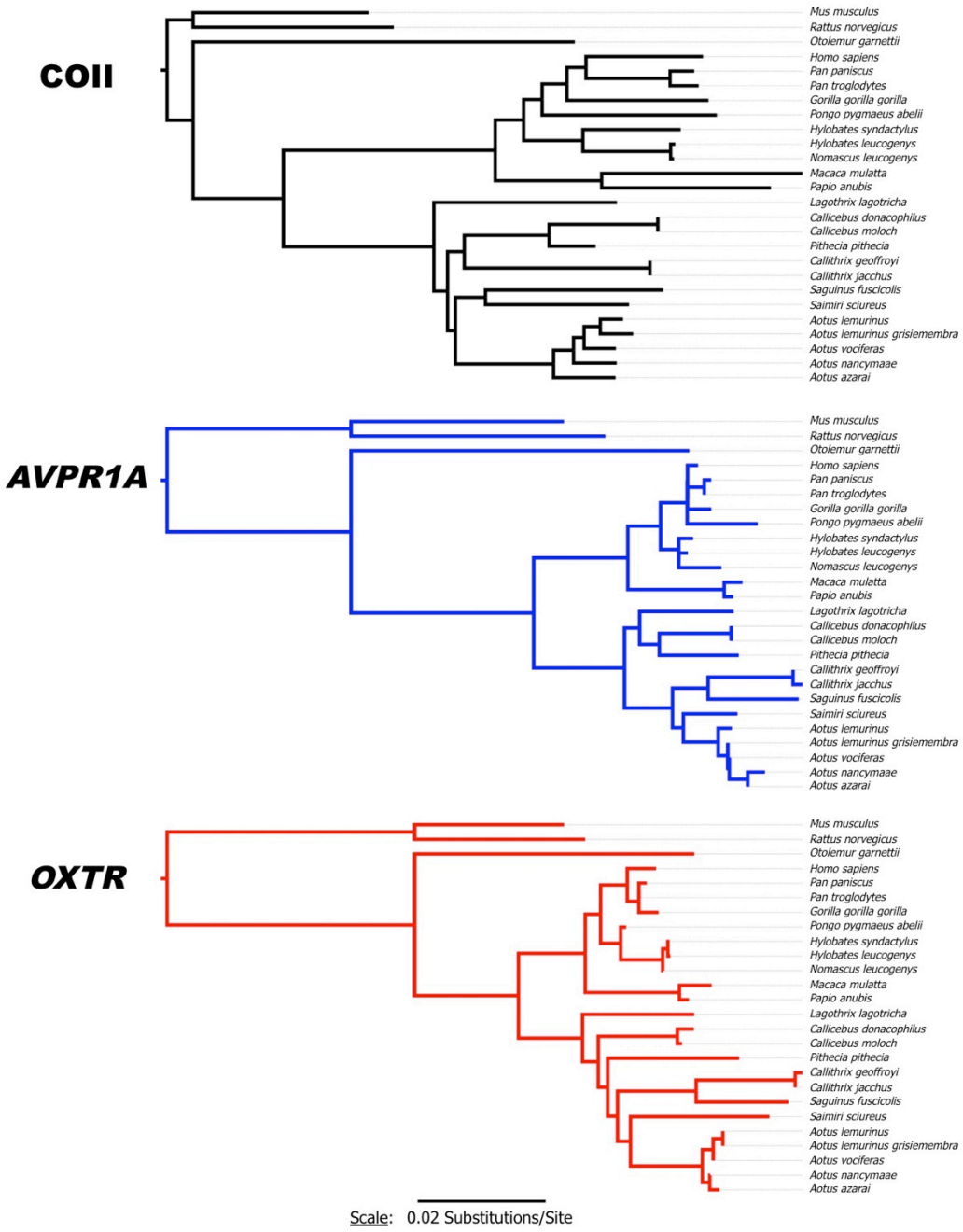


Figure 7.1. Three-way comparison of nucleotide-based phylograms for three loci in 26 primate and rodent taxa. All trees constructed using maximum likelihood criteria applying and applying the HKY nucleotide model.

The functionality of this variation was clarified through the use of a modified HKA test of non-neutrality across multiple loci (**Table 7.5**, Hudson et al., 1987). Looking at the distribution of nucleotide variance at COII, *AVPR1A*, and *OXTR* within a species (*A. azarai*), we once again note their intraspecific conservation. In our between-species comparison (*A. azarai* vs. *A. nancymaae*), a greater amount of nucleotide polymorphism was present at each locus, with the greatest number of substitutions occurring in the COII gene. At the inter-generic level (*A. azarai* vs. *Lagothrix lagotricha*), we observed even more nucleotide variation in all loci, and this variation was particularly pronounced when comparing COII (73 mutations) to *AVPR1A* (23) and *OXTR* (30). However, for COII, the bulk of these changes were functionally silent. Although *Aotus* and *Lagothrix* possess 73 nucleotide changes in 549 bp of the gene, only four constituted amino acid replacements. For *AVPR1A*, 23 nucleotide changes result in seven non-synonymous changes, while in *OXTR* the 30 observed nucleotide substitutions result in six amino acid changes.

Table 7.5. Distribution of polymorphisms in *COII*, *AVPR1A* and *OXTR* within species, and between species and genera.

	COII		***AVPR1A***		***OXTR***	
	No. sites compared	**No. variable sites**	**No. sites compared**	**No. variable sites**	**No. sites compared**	**No. variable sites**
Within species (n=25)	696	0	563	0	922	1
Between species*	696	35	563	3	922	2
Between genera†	**549**	**73 [4]**	**563**	**23 [7]**	**922**	**30 [6]**

*Between species comparison was performed using one *A. azarai* individual (AaaPLunk) and one *A. nancymaae* individual (Ana006).
†Between genera comparison was performed using one *A. azarai* individual (AaaPLunk) and one *Lagothrix lagotricha* individual (Lla001).
Bracketed ([x]) numbers indicated number of non-synonymous amino acid changes resulting from nucleotide site variation.

Although the COII gene is not a true neutral locus, it is readily apparent that both *AVPR1A* and *OXTR* have much higher rates of non-synonymous to synonymous mutations across primate genera, suggesting that positive or diversifying selection has influenced their evolution. By contrast, the high degree of intraspecific sequence conservation at *AVPR1A* and *OXTR* suggests the effects of negative or purifying selection. The combination of these results would seem to suggest that ancestral primate populations at one time exhibited greater variation in the coding regions of these genes, and that certain sequence patterns were selected (or resulted from genetic drift) and eventually conserved in the different lineages of extant primate species.

7.7 Conclusions

Primates display a host of social systems, and even closely related primate taxa exhibit different mating strategies and social behaviors. Ecological conditions may be a strong selective influence on the evolution of primate social behavior, and molecular sequence changes could serve as a means of adapting to these ecological challenges. The coding regions of genes likely associated with social behaviors, such as *AVPR1A* and *OXTR*, allow for nucleotide substitutions to potentially change the structure and function of proteins through non-synonymous amino acid changes. In contrast, regulatory regions and promoters related to such genes provide additional locations for mutations to accumulate, and, in some cases, for facilitating quick adaptive responses (via gene transcriptional regulation) to environmental change.

The Azara's owl monkeys represent a true socially and genetically monogamous primate species. Pro-social behaviors related to this mating system are shared by other

members of the genus, as well as several other New World monkey taxa, and are exhibited throughout a range of ecological contexts in Central and South America. Although some level of genetic diversity of *AVPR1A* and *OXTR* in platyrrhines is undoubtedly a result of the colonization of the New World by proto-platyrrhines some 30 million years ago (Tarling, 1980; Poux et al., 2006; Hodgson et al., 2009), it may also reflect the consequences of selection acting upon this locus, especially given the enrichment of monogamous social systems among New World species (Di Fiore and Rendell, 1994; Schwindt et al., 2004; Rendell and Di Fiore, 2007).

The evolution of primate social behavior must have a historical component related to past stochastic genetic drift and population events, a selective or adaptive component related to ecological and reproductive pressures, and a molecular mechanistic component related to gene structure, gene dosage, and overall genomic variation. Thus, there is no simple one-to-one correlation of diversity found at the *AVPR1A* and *OXTR* genes with the observed social behaviors and mating systems of owl monkeys and other primates. In other words, it is clear that more than two genetic loci must be involved in the manifestation of social behaviors, as variation at both *AVPR1A* and *OXTR* fails to sort out with the phylogenetic status of monogamous primate taxa. Clearly, there is no single "monogamy gene"; the story behind the evolution of primate sociality is much more complicated.

The New World monkeys have experienced many lineage-specific amino acid substitutions in *AVPR1A* and *OXTR* following their adaptive radiation on the continent 30 million years ago, implying that a population bottleneck and/or ancient positive selection helped to shape the variation we observe at both loci. Yet, negative selection

characterizes both loci at the species-level in many primate lineages (including owl monkeys), indicating that, following the ancient splitting of populations of proto-platyrrhines, certain behavioral regimes were initially quite diverse, then selected and molecularly conserved. In addition, the regulatory regions of *AVPR1A* and *OXTR* exhibit significant forms of variation that drives their differential expression across primate taxa, supporting the notion that differences in gene dosage, rather than just changes in protein structure, could be a large, if not primary, source of phenotypic differences between primate species (King and Wilson, 1975; Robinson et al., 2005). Thus, it is possible that the genetic changes responsible for the behavioral differences between primate species may be both quantitative (expression) and qualitative (protein structure).

In conclusion, we have characterized the molecular features of *AVPR1A* and *OXTR* in *A. azarai* and other platyrrhines, and demonstrated that there are substantial sequence differences at both loci across primate species. These data provide new clues about the possible basis of pair bonding in New World species, and may help to explain the sporadic appearance of monogamy in this infraorder. Despite a common molecular origin, we argue that the AVP and OT pathways have evolved in markedly different ways, due in part to their chromosomal locations and their relative proximity to regions of molecular instability. This study reinforces the notion that primates have experienced significant neurogenetic variation during their evolution, and suggests that monogamy has arisen multiple times in the primate order through different molecular mechanisms.

References

Abrahams BS, Geschwind DH. (2008). Advances in autism genetics: on the threshold of a new neurobiology. *Nature Reviews Genetics 9*, 341-355.

Adkins RM, Honeycutt RL. (1994). Evolution of the primate cytochrome *c* oxidase subunit II gene. *J Mol Evol 38*, 215-231.

Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. (2004). Parallel metropolis-coupled Markov Chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics 20*, 407-415.

Alvard M. (2003). The adaptive nature of culture. *Evolutionary Anthropology 12(3*), 136-149.

Amaral JM, Simoes AL, De Jong D. (2005). Allele frequencies and genetic diversity in two groups of wild tufted capuchin monkeys (*Cebus apella nigritus*) living in an urban forest fragment. *Genet Mol Res 4*, 832-838.

Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu X, Janke A. (2002). Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc Natl Acad Sci USA 99*, 8151-8156.

Ascunce MS, Hasson E, Mudry MD. (2002). Description of the cytochrome *c* oxidase subunit II gene in some genera of New World monkeys (Primates, Platyrrhini). *Genetica 114*, 253-267

Ashley MV, Vaughn JL. (1995). Owl monkeys (*Aotus*) are highly divergent in mitochondrial cytochrome *c* oxidase (COII) sequences. *Int J Primatol 16*, 793-806.

Babb PL, Fernandez-Duque E, Schurr TG. (2010). *AVPR1A* sequence variation in monogamous owl monkeys (*Aotus azarai*) and its implications for the evolution of platyrrhine social behavior. *J Mol Evol 71*, 279-297.

Babb PL, Fernandez-Duque E, Baiduc CA, Gagneux P, Evans S, Schurr TG. (2011). MtDNA diversity in Azara's owl monkeys (*Aotus azarai azarai*) of the Argentinean Chaco. *Am J Phys Anthropol 146*, 209-224.

Baker BS, Taylor BJ, Hall JC. (2001). Are complex behaviors specified by dedicated regulatory genes? Reasoning from *Drosophila*. *Cell 105*, 13-24.

Bales KL, Mason WA, Catana C, Cherry SR, Mendoza SP. (2007a). Neural correlates of pair-bonding in a monogamous primate. *Brain Res 1184*, 245-253.

Bales KL, van Westerhuyzen JA, Lewis-Reese AD, Grotte ND, Lanter JA, Carter CS. (2007b). Oxytocin has dose-dependent developmental effects on pair-bonding and alloparental care in female prairie voles. *Horm Behav 52*(2), 274-279.

Bales KL, Plotsky PM, Young LJ, Lim MM, Grotte N, Ferrer E, Carter CS. (2007c). Neonatal oxytocin manipulations have long- lasting, sexually dimorphic effects on vasopressin receptors. *Neurosci 144,* 38-45.

Bandelt HJ, Forster P, Sykes BC, Richards MB. (1995). Mitochondrial portraits of human populations. *Genetics 141*, 743-753.

Bandelt HJ, Forster P, Röhl A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol 16*, 37-48.

Bandelt HJ, Parson W. (2008). Consistent treatment of length variants in the human mtDNA control region: a reappraisal. *Int J Legal Med 122*, 11–21.

Bandelt HJ, Macaulay V, Richards M. (2000). Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol Phylogenet Evol 16*, 8–28.

Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V. (2002). The fingerprint of phantom mutations in mitochondrial DNA data. *Am J Hum Genet 71*, 1150–1160.

Beard KC, Wang J. (2004). The eosimiid primates (Anthropoidea) of the Heti Formation, Yuanqu Basin, Shanxi and Henan Provinces, People's Republic of China. *J Hum Evol 46*, 401-432.

Belich MP, Madrigal JA, Hildebrand WH, Zemmour J, Williams RC, Luz R, Petzl-Erler ML, Parham, P. (1992). Unusual HLA-B alleles in two tribes of Brazilian Indians. *Nature 357*, 326-329.

Bell SD, Magill CP, Jackson SP. (2001). Basal and regulated transcription in Archaea. *Biochem Soc Trans 29*, 392-395.

Bennett AJ, Lesch KP, Heils A, Long JC, Lorenz JG, Shoaf SE, Champoux M, Suomi SJ, Linnoila MV, Higley JD. (2002). Early experience and serotonin transporter gene variation interact to influence primate CNS function. *Molecular Psychiatry 7*, 118–122.

Bertonatti C, Corcuera J. (2000). Situación Ambiental Argentina, Fundación Vida Silvestre Argentina. Buenos Aires, Argentina.

Bloch JI, Silcox MT, Boyer DM, Sargis EJ. (2007). New Paleocene skeletons and the relationship of plesiadapiforms to crown-clade primates. *Proc Natl Acad Sci USA 104*, 1159-1164.

Böhle U-R, Zischler H. (2002). Polymorphic microsatellite loci for the mustached tamarin (*Saguinus mystax*) and their cross-species amplification in other New World monkeys. *Mol Ecol Notes 2*, 1-3.

Boinski S. (1999). The social organizations of squirrel monkeys: Implications for ecological models of social evolution. *Evolutionary Anthropology 8(3)*, 101-114.

Brown WM. (1979). Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA 76*, 1967-1971.

Burmeister M, McInnis MG, Zöllner S. (2008). Psychiatric genetics: progress amid controversy. *Nature Reviews Genetics 9*, 527-540.

Brumback RA. (1973). Two distinctive types of owl monkeys (*Aotus*). *J Med Primatol 2*, 284-289.

Brumback RA. (1974). A third species of the owl monkey (*Aotus*). *J Hered 65*, 321-323.

Brumback RA, Staton RD, Benjamin SA, Lang CM. (1971). The chromosomes of *Aotus trivirgatus* (Humboldt 1812). *Folia Primatol 15*, 264-273.

Carter CS. (1998). Neuroendocrine perspectives on social attachment and love. *Psychoneuroendocrinology 23(8)*, 779-818.

Carter CS, Grippo AJ, Pournajafi-Nazarloo H, Ruscio MG, Porges SW. (2008). Oxytocin, vasopressin and sociality. *Prog Brain Res*, *170*, 331-336.

Champoux M, Bennett A, Shannon C, Higley JD, Lesch KP, Suomi SJ. (2002). Serotonin transporter gene polymorphism, differential early rearing, and behavior in rhesus monkey neonates. *Molecular Psychiatry 7(10)*, 1058-1063.

Cheney DL, Seyfarth RM. (1999). Recognition of other individuals' social relationships by female baboons. *Animal Behaviour 58*, 67-75.

Cheney DL, Seyfarth RM, Palombit R. (1996). The function and mechanisms underlying baboon 'contact' barks. *Animal Behaviour 52*, 507-518.

Collura RV, Stewart CBR, Ruvolo M. (Unpublished). Isofunctional remodeling of anthropoid primate mitochondrial proteins. Submitted to GenBank 6 March 2003. Direct submission.

Conrad D, Pinto D, Redon R, Feuk L, Gökçümen Ö, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, MacArthur DG, MacDonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Wellcome Trust Case Control Consortium, Tyler-Smith C,

Carter NP, Lee C, Scherer SW, Hurles ME. (2010). Common copy number variation in the human genome: mechanism, selection and disease association. *Nature 464*, 704-712.

Defler TR, Bueno ML. (2003). Karyological guidelines for *Aotus* taxonomy. *Am J Primatol 60*, 134-135.

Defler TR, Bueno ML. (2007). *Aotus* diversity and the species problem. *Primate Conserv 22*, 55-70.

Defler TR, Bueno ML, Hernández-Camacho JI. (2001). The taxonomic status of *Aotus hershkovitzi*: Its relationship to *Aotus lemurinus lemurinus*. *Neotrop Primates 9*, 37-52.

De la Balze V, Biani M, Montani R. (2003). El gran Chaco Americano: un manual para acercarnos a sus componentes ambientales y sociales en la Argentina. Buenos Aires, AR: Impreso. Pp.1-127.

Demir E, Dickson BJ. (2005). *fruitless* splicing specifies male courtship behavior in *Drosophila*. *Cell 121*, 785-794.

de Waal FB. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annu Rev Psychol 59*, 279-300.

Di Fiore A, Rendell D. (1994) Evolution of social organization: A reappraisal for primates by using phylogenetic methods. *Proc Natl Acad Sci USA 91*, 9941-9945.

Di Fiore A. (2003). Molecular genetic approaches to the study of primate behavior, social organization, and reproduction. *Am J Phys Anthropol, Suppl 37,* 62-99.

Di Fiore A, Fleischer RC. (2004). Microsatellite markers for woolly monkeys (*Lagothrix lagotricha*) and their amplification in other New World primates (Primates: Platyrrhini). *Mol Ecol Notes 4*, 246-249.

Di Fiore A, Fleischer RC. (2005). Social behavior, reproductive strategies, and population genetic structure of *Lagothrix poeppigii*. *Int J Primatol 26*, 1137-1173.

Di Fiore A, Disotell T, Gagneux P, Ayala FJ. (2009). Primate malarias: evolution, adaptation, and species jumping. In *Primate Parasite Ecology: The Dynamics and Study of Host-Parasite Relationships*. Huffman MA, Chapman CA, eds. Cambridge, UK: Cambridge University Press. Pp.141-182.

Disotell TR, Honeycutt RL, Ruvolo M. (1992). Mitochondrial DNA phylogeny of the Old-World monkey tribe Papionini. *Mol Biol Evol 9*, 1-13.

Do Nascimento FF, Bonvoicino CR, Seuanez H. (2007). Population genetic studies of

Alouatta caraya (Alouattinae, Primates): Inferences on geographic distribution and ecology. *Am J Primatol 69*, 1093-1104.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. (1991). Touchdown PCR to circumvent spurious priming during gene amplification. *Nuc Acids Res 19*, 4008.

Donaldson ZR, Kondrashov FA, Putnam A, Bai Y, Stoinski TL, Hammock EA, Young LJ. (2008). Evolution of a behavior-linked microsatellite-containing element in the 5' flanking region of the primate *AVPR1A* gene. *BMC Evol Biol 23*, 180.

Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics 161*, 1307-1320.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol 4*, e88.

Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A. (2010). Geneious 5.0, available at http://www.geneious.com

Drummond AJ, Rambaut A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol 7*, 214.

Edwards DH, Heitler WJ, Krasne FB. (1999). Fifty years of a command neuron: the neurobiology of escape behavior in the crayfish. *Trends in Neuroscience 22*, 153-161.

Edwards S, Self DW. (2006). Monogamy: dopamine ties the knot. *Nat Neurosci 9,* 7-8.

Egashira N, Tanoue A, Matsuda T, Koushi E. (2007). Impaired social interaction and reduced anxiety-related behavior in vasopressin V1a receptor knockout mice. *Behav Brain Res 178,* 125.

Ellsworth JA, Hoelzer GA. (1998). Characterization of microsatellite loci in a New World primate, the mantled howler monkey (*Alouatta palliata*). *Mol Ecol 7*, 657-658.

Enard W, Pääbo S. (2004). Comparative primate genomics. *Annu Rev Genomics Hum Genet 5*, 351–378.

Escobar-Paramo P. (2000). Microsatellite primers for the wild brown capuchin monkey *Cebus apella*. *Mol Ecol 9*, 107-108.

Estoup A, Tailliez C, Cornuet JM, Solignac M. (1995). Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Mol Biol Evol 12*, 1074-1084.

Excoffier L, Laval G, Schneider S. (2005). Arlequin v3.0: An integrated software package for population genetics data analysis. *Evol Bioinfor 1*, e47-e50.

Felsenstein J. (2005). PHYLIP (Phylogeny Inference Package) v3.6, available at http://evolution.genetics.washington.edu/phylip.html

Fernandez-Duque, E. (2011). The Aotinae: social monogamy in the only nocturnal anthropoids. In *Primates in Perspective*. Campbell CJ, Fuentes A, MacKinnon KC, Bearder SK, Stumpf RM, eds. Oxford: Oxford University Press. Pp.140-154.

Fernandez-Duque E, Rotundo M. (2003). Field methods for capturing and marking Azara's night monkeys. *Int J Primatol 24*, 1113-1120.

Fernandez-Duque E, Juárez C, Di Fiore A. (2008). Adult male replacement and subsequent infant care by male and siblings. *Primates 41*, 81-84.

Fernandez-Duque E, Valeggia C, Mendoza SP. (2009). The biology of paternal care. *Ann Rev Anthropol 38*, 83-97.

Fernandez-Duque E, Rotundo M, Sloan C. (2001). Density and population structure of owl monkeys (*Aotus azarai*) in the Argentinean Chaco. *Am J Primatol 53*, 99-108.

FigTree v1.3.1. (2009). A. Rambaut, available at http://beast.bio.ed.ac.uk/FigTree.

Fink S, Excoffier L, Heckel G. (2007). High variability and non-neutral evolution of the mammalian *avpr1a* gene. *BMC Evol Biol, 7*, 176.

Fink S, Excoffier L, Heckel G. (2006). Mammalian monogamy is not controlled by a single gene. *Proc Natl Acad Sci USA 103,* 10956-10960.

Fleagle JG. (1999). *Primate Adaptation and Evolution, 2nd ed*. San Diego: Academic Press.

Ford SM. (1986). Systematics of the New World monkeys. In: *Comparative primate biology, volume I: systematics, evolution and anatomy*. Swindler DR, Erwin J, eds. New York, NY: Alan R Liss. Pp. 73-135.

Ford SM. (1994). Taxonomy and distribution of the owl monkey. In *Aotus: The owl monkey.* Baer JF, Weller RE, Kakoma I, eds. San Diego, CA: Academic Press. Pp.1-57.

Forster P, Harding R, Torroni A, Bandelt HJ. (1996). Origin and evolution of Native American mtDNA variation: A reappraisal. *Am J Hum Genet 59*, 935-945.

Fu YX. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics 147*, 915-925.

Gemmell NJ, Western PS, Watson JM, Marshall Graves JA. (1996). Evolution of the mammalian mitochondrial control region - comparisons of control region sequences between monotreme and therian mammals. *Mol Biol Evol 13*, 798-808.

Gendzekhadze K, Norman PJ, Abi-Rached L, Layrisse Z, Parham P. (2006). High KIR diversity in Amerindians is maintained using few gene-content haplotypes. *Immunogenet 58*, 474-480.

Gilad Y, Rosenberg S, Przeworski M, Lancet D, Skorecki K. (2002). Evidence for positive selection and population structure at the human MAO-A gene. *Proc Natl Acad Sci USA 99*, 862-867.

Gill KS. (1963). A mutation causing abnormal courtship and mating behavior in males of *Drosophila melanogaster*. *American Zoologist 3*, 507.

Gökçümen Ö, Babb PL, Iskow R, Zhu Q, Shi, X, Mills RE, Ionita-Laza I, Vallender EJ, Clark AG, Lee C, Johnson WE. (2011). Refinement of primate copy number variation hotspots identifies candidate genomic regions evolving under positive selection. *Genome Biology 12*, R52.

Gökçümen Ö, Lee C. (2009). Copy number variants (CNVs) in primate species using array-based comparative genomic hybridization. *Methods 49*, 18-25.

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves P. (1998). Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol Phylogenet Evol 9(3),* 585-598.

Grativol AD, Ballou JD, Fleischer RC. (2001). Microsatellite variation within and among recently fragmented populations of the golden lion tamarin (*Leontopithecus rosalia*). *Conserv Genet 2*, 1-9.

Gregory SG, Connelly JJ, Towers AJ, Johnson J, Biscocho D, Markunas CA, Lintas C, Abramson RK, Wright HH, Ellis P, Langford CF, Worley G, Delong GR, Murphy SK, Cuccaro ML, Persico A, Pericak-Vance MA. (2009). Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. *BMC Medicine, 7,* e1-e13.

Grippo AJ, Wu KD, Hassan I. (2008). Social isolation in prairie voles induces behaviors relevant to negative affect: toward the development of a rodent model focused on co-occurring depression and anxiety. *Depress Anxiety 25,* e17-26.

Grippo, A J, Trahanas DM, Zimmerman RRN, Porges SW, Carter CS. (2009). Oxytocin protects against negative behavioral and autonomic consequences of long-term social isolation. *Psychoneuroendocrinology, 34(10),* 1542-1553.

Guindon S, Gascuel O. (2003). A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Sys Biol 52*, 696-704.

Gwee P-C, Amemiya CT, Brenner S, Venkatesh B. (2008). Sequence and organization of coelacanth neurohypophysial hormone genes: evolutionary history of the vertebrate neurohypophysial hormone gene locus. *BMC Evol Biol 93*, 1-12.

Hamilton WD. (1964a). The genetical evolution of social behaviour. I. *J Theoret Biol 7*, 1–16.

Hamilton WD. (1964b). The genetical evolution of social behaviour. II. *J Theoret Biol 7,* 17–52.

Hamilton MB, Pincus EL, Di Fiore A, Fleischer RC. (1999). Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques 27*, 500-7.

Hammock EA, Lim MM, Nair HP, Young LJ. (2005). Association of vasopressin 1a receptor levels with a regulatory microsatellite and behavior. *Genes Brain Behav 4,* 289-301.

Hammock E, Young L. (2006). Oxytocin, vasopressin and pair bonding: implications for autism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*, 2187.

Hammock EA, Young LJ. (2005). Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science 308(5728)*, 1630-1634.

Hammock EA, Young LJ. (2004). Functional microsatellite polymorphism associated with divergent social structure in vole species. *Mol Biol Evol 21(6)*, 1057-1063.

Hammock EA, Young LJ. (2002). Variation in the vasopressin V1a receptor promoter and expression: implications for inter- and intraspecific variation in social behaviour. *Eur J Neurosci 16(3)*, 399-402.

Hartwig WC, Meldrum DJ. (2002). Miocene platyrrhines of the northern Neotropics. In *The Primate Fossil Record*. WC Hartwig, ed. Cambridge, UK: Cambridge University Press. Pp.175-188.

Hartwig W. (2007). Primate Evolution. In *Primates in Perspective*. Campbell CJ, Fuentes A, MacKinnon KC, Panger M, Bearder SK. New York: Oxford University Press. Pp. 11-22.

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.

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.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ. (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol 22*, 1561-1568.

Hodgson JA, Sterner KN, Matthews LJ, Burrell AS, Jani RA, Raaum RL, Stewart CB, Disotell TR. (2009). Successive radiations, not stasis, in the South American primate fauna. *Proc Natl Acad Sci USA, 106(14)*, 5534-5539.

Hong KW, Matsukawa R, Hirata Y, Hayasaka I, Murayama Y, Ito S, Inoue-Murayama M. (2009). Allele distribution and effect on reporter gene expression of vasopressin receptor gene (AVPR1a)-linked VNTR in primates. *J Neural Transm 116*, 535-538.

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. *Mol Phylogenet Evol 4(4)*, 448-456.

Houle A. (1999). The origin of platyrrhines: An evaluation of the Antarctic scenario and the floating island model. *Am J Phys Anthropol 109,* 541-559

Huck M, Lottker P, Böhle UR, Heymann EW. (2005). Paternity and kinship patterns in polyandrous moustached tamarins (*Saguinus mystax*). *Am J Phys Anthropol 127*, 449-464.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science 294*, 2310-2314.

Huelsenbeck JP, Ronquist F. (2001). MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics 17*, 754-755.

Hudson RR, Kreitman M, Aguadé M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics 116*, 153-159.

Hugot, J-P. (1998). Phylogeny of Neotropical Monkeys: The Interplay of Morphological, Molecular, and Parasitological Data. *Molecular Phylogenetics and Evolution 9(3)*, 408-413.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. (2004). Detection of large-scale variation in the human genome. *Nat Genet 36*, 949-951.

Inoue H, Hiroshi Nojima H, Okayama H. (1990). High efficiency transformation of Escherichia coli with plasmids. *Gene 96*, 23-28.

Inoue T, Kimura T, Azuma C, Inazawa J, Takemura M, Kikuchi T, Kubota Y, Ogita K, Saji F. (1994). Structural organization of the human oxytocin receptor gene. *Journal of Biological Chemistry 269,* 32451-6.

Insel TR. (2003). Is social attachment an addictive disorder? *Physiol Behav 79*, 351-357.

Iriondo MH. (1984). The Quaternary of Northeastern Argentina. *Quaternary of South America and Peninsula Antarctica 2*, 51-78.

Iriondo MH. (1993). Geomorphology and Late Quaternary of the Chaco (South-America). *Geomorphology 7*, 289-303.

Karolchik D, Kuhn RM, Baertsch R, Barber GP, Clawson H, Diekhans M, Giardine B, Harte RA, Hinrichs AS, Hsu F, Miller W, Pedersen JS, Pohl A, Raney BJ, Rhead B, Rosenbloom KR, Smith KE, Stanke M, Thakkapallayil A, Trumbower H, Wang T, Zweig AS, Haussler D, Kent WJ. (2008). The UCSC Genome Browser Database: 2008 update. *Nuc Acids Res 36*, 773-779.

Katoh H, Takabayashi S, Itoh T. (2009). Development of microsatellite DNA markers and their chromosome assignment in the common marmoset. *Am J Primatol 71*, 912-918.

Kay RF. (1990). The phyletic relationships of extant and fossil Pitheciinae (Platyrrhini, Anthropoidea). *J Hum Evol 19*, 175-208.

Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA, Sajantila A, Tyler-Smith C. (2004). A comprehensive survey of human Y-chromosomal microsatellites. *Am J Hum Genet 74*, 1183-1197.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. (2002). The human genome browser at UCSC. *Genome Res 12*, 996-1006.

Kim SJ, Young LJ, Gonen D, Veenstra-VanderWeele J, Courchesne R, Courchesne E, Lord C, Leventhal BL, Cook EH Jr, Insel TR. (2002). Transmission disequilibrium testing of arginine vasopressin receptor 1A (*AVPR1A*) polymorphisms in autism. *Mol Psychiatry 7*, 503-507.

Kimura M. (1981). Possibility of extensive neutral evolution under stabilizing selection

with special reference to nonrandom usage of synonymous codons. *Proc Natl Acad Sci USA 78*, 5773-5777.

King MC, Wilson EO. (1975). Evolution at two levels in humans and chimpanzees. *Science 11*, 107-116.

Kleiman DG. (1977). Monogamy in mammals. *Quart Rev Biol 52*, 39-69.

Korbie DJ, Mattick JS. (2008). Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat Protocols 3*, 1452-1456.

Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A. (2008). Patterns of positive selection in six mammalian genomes. *PLoS Genet 4(8)*, e1000144.

Lau J, Fernandez-Duque E, Evans S, Dixson AF, Ryder O. (2004). Heterologous amplification and diversity of microsatellite loci in three owl monkey species (*Aotus azarai, A. lemurinus, A. nancymaae*). *Conserv Genet 5*, 727-731.

Lavocat R. (1993). Conclusions. In: *The Africa-South America Connection*. George W, Lavocat R, eds. Oxford: Clarendon Press. Pp.142-150.

Lee AG, Cool DR, Grunwald WC, Neal DE, Buckmaster CL, Cheng MY, Hyde SA, Lyons DM, Parker KJ. (2011). A novel form of oxytocin in New World monkeys. *Biol Letters 7*, 584-587.

Lee AS, Gutiérrez-Arcelus M, Perry GH, Vallender EJ, Johnson WE, Miller GM, Korbel JO, Lee C. (2008). Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies. *Hum Mol Genet 17(8)*, 1127-1136.

Leonard, CM, Rolls ET, Wilson FA, & Baylis GC. (1985). Neurons in the amygdala of the monkey with responses selective for faces. *Behav Brain Res 15(2)*, 159-176.

Lesch KP, Bengel D, Heils A, Sabol SZ, Greenberg BD, Petri S, Benjamin J, Muller CR, Hamer DH, Murphy DL. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science 274*, 1527-1531.

Lim MM, Hammock EA, Young LJ. (2004a). The role of vasopressin in the genetic and neural regulation of monogamy. *Neuroendocrin 16*, 325-332.

Lim MM, Wang Z, Olazábal DE, Ren X, Terwilliger EF, Young LJ. (2004b). Enhanced partner preference in a promiscuous species by manipulating the expression of a single gene. *Nature 429*, 754-757.

Ma NSF. (1981). Chromosome evolution in the owl monkey, *Aotus*. *Am J Phys Anthropol 54*, 293-303.

Ma NSF. (1983). Gene map of the new world Bolivian owl monkey, *Aotus*. *J Hered 74*, 27-33.

Ma NSF, Aquino R, Collins WE. (1985). Two new karyotypes in the Peruvian owl monkey (*Aotus trivirgatus*). *Am J Primatol 9*, 333-341.

Ma NSF, Elliot MW, Morgan LM, Miller AC, Jones TC. (1976a). Translocation of Y chromosome to an autosome in the Bolivian owl monkey, *Aotus*. *Am J Phys Anthropol 45*, 191-202.

Ma NSF, Jones TC, Bedard MT, Miller AC, Morgan LM, Adams EA. (1977). The chromosome complement of an *Aotus* hybrid. *J Hered 68*, 409-412.

Ma NSF, Jones TC, Miller A, Morgan L, Adams E. (1976b). Chromosome polymorphism and banding patterns in the owl monkey (*Aotus*). *Lab Animal Sci 26*, 1022-1036.

Ma NSF, Rossan RN, Kelley ST, Harper JS, Bedard MT, Jones TC. (1978). Banding patterns of the chromosomes of two new karyotypes of the owl monkey, *Aotus*, captured in Panama. *J Med Primatol 7*, 146–155.

MacKenzie MM, McGrew WC, Chamove AS. (1985). Social preferences in stump-tailed macaques (*Macaca arctoides*): effects of companionship, kinship, and rearing. *Dev Psychobiol 18(2)*, 115-123.

Maddison DR, Maddison WP. (2003). MacClade 4: Analysis of phylogeny and character evolution, v4.06. Sunderland, MA: Sinauer Associates.

Maestripieri D, Higley JD, Lindell SG, Newman TK, McCormack KM, Sanchez MM. (2006a). Early maternal rejection affects the development of monoaminergic systems and adult abusive parenting in rhesus macaques (*Macaca mulatta*). *Behav Neurosci 120(5)*, 1017-1024.

Maestripieri D, McCormack K, Lindell SG, Higley JD, Sanchez MM. (2006b). Influence of parenting style on the offspring's behaviour and CSF monoamine metabolite levels in crossfostered and noncrossfostered female rhesus macaques. *Behav Brain Res 175(1)*, 90-95.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. (2009). Finding the missing heritability of complex diseases. *Nature 461(7265)*, 747-753.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol 7*, 639-655.

Mayeaux DJ, Mason WA, Mendoza SP. (2002). Developmental changes in responsiveness to parents and unfamiliar adults in a monogamous monkey (*Callicebus moloch*). *Am J Primatol 58(2)*, 71-89.

McDonald JH, Kreitman M. (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature 351*, 652-654.

Mendoza SP, Reeder DM, Mason WA. (2002). Nature of proximate mechanisms underlying primate social systems: Simplicity and redundancy. *Evol Anthropol 11*, 112-116.

Menezes AN, Bonvicino CR, Seuanez HN. (2010). Identification, classification and evolution of owl monkeys (*Aotus*, Illiger 1811). *BMC Evol Biol 10*, 248-263.

Meyer M, Stenzel U, Myles S, Prufer K, Hofreiter M. (2007). Targeted high-throughput sequencing of tagged nucleic acid samples. *Nuc Acids Res 35*, e97.

Mock DW, Fujioka M. (1990). Monogamy and long-term pair bonding in vertebrates. *Trends Ecol Evol 5*, 39-43.

Moller AP. (2003). The evolution of monogamy: mating relationships, parental care and sexual selection. In *Monogamy: Mating Strategies and Partnerships in Birds, Humans and Other Mammals*. Reichard UH, Boesch C, eds. Cambridge: University of Cambridge Press. Pp.29-41.

Monsalve MV, Defler TR. (2011). An *Aotus* karyotype from extreme eastern Colombia. *Primate Conserv 25*, e1-e6. Electronic publication prior to print.

Morin PA, Kanthaswamy S, Smith DG. (1997). Simple sequence repeat (SSR) polymorphisms for colony management and population genetics in rhesus macaques (*Macaca mulatta*). *Am J Primatol 42*, 199-213.

Morin PA, Mahboubi P, Wedel S, Rogers J. (1998). Rapid screening and comparison of human microsatellite markers in baboons: allele size is conserved, but allele number is not. *Genomics 53*, 12-20.

Morin PA, Chambers KE, Boesch C, Vigilant L. (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Mol Ecol 10*, 1835-1844.

Muniz L, Vigilant L. (2008). Isolation and characterization of microsatellite markers in

the white-faced capuchin monkey (*Cebus capucinus*) and cross-species amplification in other New World monkeys. *Mol Ecol Resources 8*, 402-405.

Murasawa S, Matsubara H, Kijima K, Maruyama K, Mori Y, Inada M. (1995). Structure of the rat V1a vasopressin receptor gene and characterization of its promoter region and complete cDNA sequence of the 3'-end. *J Biol Chem 270*, 20042-20050.

Newman, TK, Syagailo YV, Barr CS, Wendland JR, Champoux M, Graessle M, Suomi SJ, Higley JD, Lesch KP. (2005). Monoamine oxidase A gene promoter variation and rearing experience influences aggressive behavior in rhesus monkeys. *Biological Psychiatry 57(2)*, 167-172.

Nievergelt CM, Mundy NI, Woodruff DS. (1998). Microsatellite primers for genotyping common marmosets (*Callithrix jacchus*) and other callitrichids. *Mol Ecol 7*, 1432-1434.

Nino-Vasquez JJ, Vogel D, Rodriguez R, Moreno A, Patarroyo ME, Pluschke G, Daubenberger CA. (2000). Sequence and diversity of DRB genes of *Aotus nancymaae*, a primate model for human malaria parasites. *Immunogenet 51*, 219-230.

Oklander LI, Zunino GE, Di Fiore A, Corach D. (2007). Isolation, characterization and evaluation of 11 autosomal STRs suitable for population studies in black and gold howler monkeys *Alouatta caraya*. *Mol Ecol Notes 7*, 117-120.

Opazo JC, Wildman DE, Prychitko T, Johnson RM, Goodman M. (2006). Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates). *Mol Phylogenet Evol 40*, 274-280.

Palmatier MA, Kang AM, Kidd KK. (1999). Global variation in the frequencies of functionally different catechol-O-methyltransferase alleles. *Biol Psychiatry, 46(4)*, 57-67.

Palombit RA. (1994). Dynamic pair bonds in hylobatids: implications regarding monogamous social systems. *Behaviour 128*, 65-101.

Palombit RA. (1999). Infanticide and the evolution of pair bonds in nonhuman primates. *Evolutionary Anthropology 7(4)*, 117-129.

Parham P, Arnett KL, Adams EJ, Little AM, Tees K, Barber LD, Marsh SG, Ohta T, Markow T, Petzl-Erler ML. (1997). Episodic evolution and turnover of HLA-B in the indigenous human populations of the Americas. *Tissue Antigens 50*, 219-232.

Park H, Kim JI, Ju YS, Gokcumen O, Mills RE, Kim S, Lee S, Suh D, Hong D, Kang HP, Yoo YJ, Shin JY, Kim HJ, Yavartanoo M, Chang YW, Ha JS, Chong W,

Hwang GR, Darvishi K, Kim H, Yang SJ, Yang KS, Kim H, Hurles ME, Scherer SW, Carter NP, Tyler-Smith C, Lee C, Seo JS. (2010). Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. *Nat Genet 42*, 400-405.

Pena SD, Chakraborty R. (1994). Paternity testing in the DNA era. *Trends Genet 10*, 204-209.

Perez-Sweeney BM, Valladares-Padua C, Burrell AS, Di Fiore A, Satkoski J, Van Coeverden de Groot PJ, Boag PT, Melnick DJ. (2005). Dinucleotide microsatellite primers designed for a critically endangered primate, the black lion tamarin (*Leontopithecus chrysopygus*). *Mol Ecol Notes 5*, 198-201.

Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, Hyland C, Stone AC, Hurles ME, Tyler-Smith C, Eichler EE, Carter NP, Lee C, Redon R. (2008). Copy number variation and evolution in humans and chimpanzees. *Genome Research 18,* 1698-1710.

Phelps SM, Campbell P, Zheng DJ, Ophir AG (2009) Beating the boojum: comparative approaches to the neurobiology of social behavior. *Neuropharma 58(1)*, 17-28.

Pieczarka JC, de Souza Barros RM, de Faria Jr. FM, Nagamachi CY. (1993). *Aotus* from the southwestern Amazon region is geographically and chromosomally intermediate between *A. azarae boliviensis* and *A. infulatus*. *Primates 34*, 197-204.

Pieczarka JC, Nagamachi CY, Muniz JA, Barros RM, Mattevi MS. (1998). Analysis of constitutive heterochromatin of *Aotus* (Cebidae, Primates) by restriction enzyme and fluorochrome bands. *Chrom Res 6*, 77-83.

Pitkow LJ, Sharer CA, Ren X, Insel TR, Terwilliger EF, Young LJ. (2001). Facilitation of affiliation and pair-bond formation by vasopressin receptor gene transfer into the ventral forebrain of a monogamous vole. *J Neurosci 21*, 7392-7396.

Plautz HL, Goncalves EC, Ferrari SF, Schneider MPC, Silva AL. (2009). Evolutionary inferences of the genus *Aotus* (Platyrrhini, Cebidae) from cytochrome c oxidase subunit II gene sequences. *Mol Phylogenet Evol 51*, 382-387.

Pollick AS, de Waal FB. (2007). Ape gestures and language evolution. *Proc Natl Acad Sci USA 104(19)*, 8184-8189.

Porzecanski AL, Cracraft J. (2005). Cladistic analysis of distributions and endemism (CADE): using raw distributions of birds to unravel the biogeography of the South American aridlands. *J Biogeog 32*, 261-275.

Posada D. (2008). jModelTest: Phylogenetic model averaging. *Mol Biol Evol 25*, 1253-1256.

Poux C, Chevret P, Huchon D, De Jong WW, Douzery EJP. (2006). Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. *Systems Biology 55(2),* 228-244.

Pusey A, Williams J, Goodall J. (1997). The influence of dominance rank on the reproductive success of female chimpanzees. *Science 277(5327*), 828-831.

Purvis A, Bromham L. (1997). Estimating the transition/transversion ratio from independent pairwise comparisons with an assumed phylogeny. *J Mol Evol 44*, 112-119.

Raaum RL, Sterner KN, Noviello CM, Stewart CB, Disotell TR. (2005). Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. *J Hum Evol 48*, 237-257.

Rambaut A, Drummond AJ. (2007). Tracer v1.5, at http://beast.bio.ed.ac.uk/Tracer

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

Ray DA, Xing J, Heges DJ, Hall MA, Laborde ME, Anders BA, White BR, Stoilova N, Fowlkes JD, Landry KE, Chemnick LG, Ryder OA, Batzer MA. (2005). *Alu* insertion loci and platyrrhine primate phylogeny. *Mol Phylogenet Evol 35,* 117-126.

Rendell D, Di Fiore A (2007) Homoplasy, homology, and the perceived special status of behavior in evolution. *J Hum Evol 52,* 504-521.

Robinson GE, Grozinger CM, & Whitfield CW. (2005). Sociogenomics: social life in molecular terms. *Nature Reviews Genetics 6(4)*, 257-270.

Roeder AD, Bonhomme M, Heijmans C, Bruford MC, Crouau-Roy B, Doxiadis G, Otting N, and the INPRIMAT consortium. (2009). A large panel of microsatellite markers for genetic studies in the infra-order Catarrhini. *Folia Primatol 80*, 63-69.

Rogers AR, Harpending HC. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol 9*, 552-569.

Ronquist F, Huelsenbeck, JP. (2003). MrBayes 3, Bayesian phylogenetic inference under mixed models. *Bioinformatics 19*, 1572-1574.

Rosenberger AL. (1984). Fossil new world monkeys dispute the molecular clock. *J Hum Evol 13*, 737-742.

Rosenberger AL, Tejedor MF, Cooke SB, Pekar S. (2009). Platyrrhine ecophylogenetics in space and time. In *South American Primates: Comparative Perspectives in the Study of Behavior, Ecology and Conservation*. Garber PA, Estrada A, Bicca-Marques JC, Heymann EW, Strier KB, eds. New York: Springer Science. Pp.69-116.

Rosso L, Keller L, Kaessmann H, Hammond RL. (2008). Mating system and *avpr1a* promoter variation in primates. *Biol Lett 4*, 375-8.

Rotundo M, Fernandez-Duque E, Dixson A. (2005). Infant development and parental care in free-ranging *Aotus azarai azarai* in Argentina. *Int J Primatol 26*, 1459-1473.

Rotundo M, Sloan C, Fernandez-Duque E. (2000). Cambios estacionales en el ritmo de actividad del mono mirikiná (*Aotus azarai*) en Formosa Argentina. In *Manejo de Fauna Silvestre en Amazonía y Latinoamérica*. Cabrera E, Mércolli C, Resquin R, eds. Asunción: Fundación Moisés Bertoni. Pp.413-417.

Rozas J, Sanchez-Delbarrio JC, Messeguer X, Rozas R. (2003). DnaSP: DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics 19*, 2496-2497.

Rozen S, Skaletsky H. (2000). Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Krawetz S, Misener S, eds. Totowa, NJ: Humana Press. Pp.365-386.

Ruvolo M, Zehr S, Von Dornum M, Pan D, Chang B, Lin J. (1993). Mitochondrial COII sequences and modern human origins. *Mol Biol Evol 10*, 1115-1135.

Sabol SZ, Hu S, Hamer D. (1998). A functional polymorphism in the monoamine oxidase A gene promoter. *Hum Genet 103(3)*, 273-279.

Schneider H. (2000). The current status of the New World monkey phylogeny. *An Acad Bras Ci 72*, 165-172.

Schwindt DM, Carillo GA, Bravo JJ, Di Fiore A, Fernandez-Duque E. (2004). Comparative socioecology of monogamous primates in the Amazon and Gran Chaco. *Folia Primatologica 75*, 412.

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Leotta A, Pai D, Zhang R, Lee YH, Hicks J, Spence SJ, Lee AT, Puura K, Lehtimäki T, Ledbetter D, Gregersen PK, Bregman J, Sutcliffe JS, Jobanputra V, Chung W, Warburton D, King MC, Skuse D, Geschwind DH,

Gilliam TC, Ye K, Wigler M. (2007). Strong Association of *de novo* copy number mutations with Autism. *Science 316*, 445-449.

Setoguchi T, Rosenberger AL. (1987). A fossil owl monkey from La Venta, Colombia. *Nature 326*, 692-694.

Sigurgardóttir S, Helgason A, Gulcher JR, Stefansson K, Donnelly PJ. (2000). The mutation rate in the human mtDNA control region. *Am J Hum Genet 66*, 1599-1609.

Simons EL, Rasmussen DT, Gebo DL. (1987). A new species of *Propliopithecus* from the Fayum, Egypt. *Am J Phys Anthropol 73*, 139-147.

Skinner B. (1984). The evolution of behavior. *J Exp Anal Behav 41*, 217-221.

Slate J, Marshall TC, Pemberton JM. (2000). A retrospective assessment of the accuracy of the paternity inference program CERVUS. *Mol Ecol 9*, 801-808.

Soulsbury CD, Iossa G, Edwards KJ. (2009). The influence of evolutionary distance between cross-species microsatellites and primer base-pair composition on allelic dropout rates. *Conserv Genet 10*, 797-802.

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

Sterner KN, Raaum RL, Zhang YP, Stewart CB, Disotell TR. (2006). Mitochondrial data support an odd-nosed colobine clade. *Mol Phylogenet Evol 40*, 1-7.

Stockinger P, Kvitsiani D, Rotkopf S, Tirian L, Dickson BJ. (2005). Neural circuitry that governs *Drosophila* male courtship behavior. *Cell 121*, 795-807.

Suarez CF, Ripoll V, Pardo L, Patarroyo MA, Patarroyo ME, Corredor V. (Unpublished). Phylogenetic analysis of the owl monkey genus *Aotus* based on the mitochondrial cytochrome oxidase II gene. Submitted to GenBank 21 February 2001. Direct submission.

Swofford DL. (2002). PAUP*: Phylogenetic analysis using parsimony (and other methods), v4.0b10 Beta. Sunderland, MA: Sinauer Associates.

Tajima F. (1989a). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics 123*, 585-595.

Tajima F. (1989b). The effect of change in population size on DNA polymorphism. *Genetics 123*, 597-601.

Takai M, Anaya F, Shigehara N, Setoguchi T. (2000). New fossil materials of the earliest new world monkey, *Branisella boliviana*, and the problem of platyrrhine origins. *Am J Phys Anthropol 111*, 263-281.

Takai M, Anaya F. (1996). New specimens of the oldest fossil platyrrhine, *Branisella boliviana*, from Salla, Bolivia. *Am J Phys Anthropol 99*, 301-317.

Takai M, Nishimura T, Shigehara N, Setoguchi T. (2009). Meaning of the canine sexual dimorphism in fossil owl monkey, *Aotus dindensis* from the middle Miocene of La Venta, Colombia. *Front Oral Biol 13*, 55-59.

Takayanagi Y, Yoshida M, Bielsky IF, Ross HE, Kawamata M, Onaka T, Yanagisawa T, Kimura T, Matzuk MM, Young LJ, Nishimori K. (2005). Pervasive social deficits, but normal parturition, in oxytocin receptor-deficient mice. *Proc Natl Acad Sci U S A 102*, 16096-16101.

Tansey KE, Hill MJ, Cochrane LE, Gill M, Anney RJL, Gallagher L. (2011). Functionality of promoter microsatellites of arginine vasopressin receptor 1a (AVPR1A): implications for autism. *Molecular Autism 2*, 1-8.

Tarling DH. (1980). The geologic evolution of South America with special reference to the last 200 million years. In *Evolutionary Biology of the New World Monkeys and Continental Drift*. Ciochon RL, Chiarelli AB, eds. New York: Plenum Press. Pp.1-42.

Tejedor MF, Tauber AA, Rosenberger AL, Swisher CC, Palacios ME. (2006). New primate genus from the Miocene of Argentina. *Proc Natl Acad Sci USA 14*, 5437-5441.

Thalmann O, Serre D, Hofreiter M, Lukas D, Eriksson J, Vigilant L. (2005). Nuclear insertions help and hinder inference of the evolutionary history of gorilla mtDNA. *Mol Ecol 14*, 179-188.

Thibonnier M, Graves MK, Wagner MS, Auzan C, Clauser E, Willard HF. (1996). Structure, sequence, expression and chromosomal location of the human V1a vasopressin receptor gene. *Genomics 31*, 327-334.

Thibonnier M, Graves MK, Wagner MS, Chatelain N, Soubrier F, Corvol P, Willard HF, Jeunemaitre X. (2000). Study of V(1)-vascular vasopressin receptor gene microsatellite polymorphisms in human essential hypertension. *J Mol Cell Cardiol 32*, 557–564.

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

Trivers RL. (1972). Parental investment and sexual selection. In: *Sexual Selection and the Descent of Man*. Campbell B, ed. Chicago: Aldine Publishing Company. Pp.136-179.

Turner LM, Young AR, Rompler H, Schoneberg T, Phelps SM, Hoekstra HE. (2010). Monogamy evolves through multiple mechanisms: Evidence from V1aR in deer mice. *Mol Biol Evol 27*, 1269-1278.

Van Schaik CP, Dunbar RIM. (1990). The evolution of monogamy in large primates: A new hypothesis and some crucial tests. *Behaviour 115,* 30-61.

Van Schaik CP, Van Hoof JARAM. (1983). On the ultimate causes of primate social systems. *Behaviour 85*, 91-117.

Van Schaik CP, Kappeler PM. (2003). The evolution of social monogamy in primates. In: *Monogamy: Mating Strategies and Partnerships in Birds, Humans and Other Mammals*. Reichard UH, Boesch C, eds. Cambridge: University of Cambridge Press. Pp. 59-80.

Van Valen L. (1973). A new evolutionary law. *Evol Theory 1*, 1-30.

Viard F, Franck P, Dubois MP, Estoup A, Jarne P. (1998). Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *J Mol Evol 47*, 42-51.

Visscher PM, Hill WG, Wray NR. (2008). Heritability in the genomics era: concepts and misconceptions. *Nature Reviews Genetics 9*, 255-266.

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(3)*, 459-476.

Walum H, Westberg L, Henningsson S, Neiderhiser JM, Reiss D, Igl W, Ganiban JM, Spotts EL, Pedersen NL, Eriksson E, Lichtenstein P. (2008). Genetic variation in the vasopressin receptor 1a gene (*AVPR1A*) associates with pair-bonding behavior in humans. *Proc Natl Acad Sci USA 105*, 14153-14156.

Weinreich DM. (2000). The rates of molecular evolution in rodent and primate mitochondrial DNA. *J Mol Evol 52*, 40-50.

Wildman DE, Jameson NM, Opazon JC, Yi SV. (2009). A fully resolved genus level phylogeny of neotropical primates (Platyrrhini). *Mol Phy Evol 53*, 694-702.

Wilson EO. (1975). *Sociobiology: the New Synthesis*. Cambridge: Harvard University Press.

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

Wolovich CK, Feged A, Evans S, Green SM. (2006). Social patterns of food sharing in monogamous owl monkeys. *Am J Primatol 68*, 663-674.

Wolovich CK, Perea-Rodriguez JP, Fernandez-Duque E. (2008). Food transfers to young and mates in wild owl monkeys (*Aotus azarai*). *Am J Primatol 70*, 211-21.

Wright PC. (1994). The behavior and ecology of the owl monkey. In: *Aotus: The Owl Monkey*. Baer JF, Weller RE, Kakoma I, eds. San Diego: Academic Press. Pp. 97-113.

Xu X, Arnason U. (1996). A complete sequence of the mitochondrial genome of the western lowland gorilla. *Mol Biol Evol 13*, 691-698.

Yang Z, Yoder AD. (1999). Estimation of the transition/transversion rate bias and species sampling. *J Mol Evol 48*, 274-283.

Yoder AD, Cartmill M, Ruvolo M, Smith K, Vigalys R. (1996). Ancient single origin for Malagasy primates. *Proc Natl Acad Sci USA 93*, 5122-5126.

Young LJ, Nilsen R, Waymire KG, MacGregor GR, Insel TR. (1999). Increased affiliative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. *Nature 400*, 766-768.

Young LJ, Murphy Young AZ, Hammock EA. (2005). Anatomy and neurochemistry of the pair bond. *J Comp Neurol 493*, 51-57.

Young LJ, Hammock EA. (2007). On switches and knobs, microsatellites and monogamy. *Trends Genet 23*, 209-212.

Zunino GE, Galliari CA, Colillas OJ. (1985). Distribución y conservación del mirikiná (*Aotus azarae*), en Argentina, resultados preliminares. In *Primatologia*. M T de Mello, ed. A Primatologia No Brasil, Campinas. Pp.305-316.

Zunino GE, Kowalewski MM, Oklander LI, Gonzalez V. (2007). Habitat fragmentation and population size of the black and gold howler monkey (*Alouatta caraya*) in a semideciduous forest in northern Argentina. *Am J Primatol 69*, 966-975.

Zwickl DJ, Hillis DM. (2002). Increased taxon sampling greatly reduces phylogenetic error. *Syst Biol 51*, 588-598.