J Mol Evol (2010) 71:279–297
DOI 10.1007/s00239-010-9383-6

AVPR1A Sequence Variation in Monogamous Owl Monkeys (*Aotus azarai*) and Its Implications for the Evolution of Platyrrhine Social Behavior

Paul L. Babb · Eduardo Fernandez-Duque · Theodore G. Schurr

Received: 29 December 2009 / Accepted: 17 August 2010 / Published online: 14 September 2010

Abstract 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 a wild population from the Gran Chaco. We also assessed the interspecific variation of *AVPR1A* in platyrrhine 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.

P. L. Babb · E. Fernandez-Duque · T. G. Schurr (✉)
Department of Anthropology, University of Pennsylvania, 344 University Museum, 3260 South Street, Philadelphia, PA 19104-6398, USA
e-mail: tgschurr@sas.upenn.edu

E. Fernandez-Duque
CECOAL-Conicet, Corrientes, Argentina

Keywords Behavioral genetics · Monogamy · V1aR · Vasopressin · Primates

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 Shaik 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*) 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 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), whereas a separate study 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 et al. 2005; Young and Hammock 2007). Overall, these studies demonstrate that the distribution of V1aR in the brain directly influences social behavior.

They further suggest that the evolution of regulatory sequences drives the expression of the gene, and may be responsible for emergence of different social structures in different species. However, STR patterns are not the only component responsible for differences in brain expression of *AVPR1A*. As noted by Fink et al. (2006), this pattern is 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 (Fig. 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 have been no studies of *AVPR1A* sequence variation in these species.

For this reason, 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 initially sequenced the entire *AVPR1A* locus in a wild population of monogamous owl monkeys (*Aotus azarai*) to characterize the genetic variation present in the 5′ promoter and coding region at the intraspecific level. We then sequenced *AVPR1A* in several platyrrhine 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 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 may have done so through different molecular mechanisms.

Methods

Study Population

Azara's owl monkeys, *A. azarai*, 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. 2005; Fernandez-Duque 2009). 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

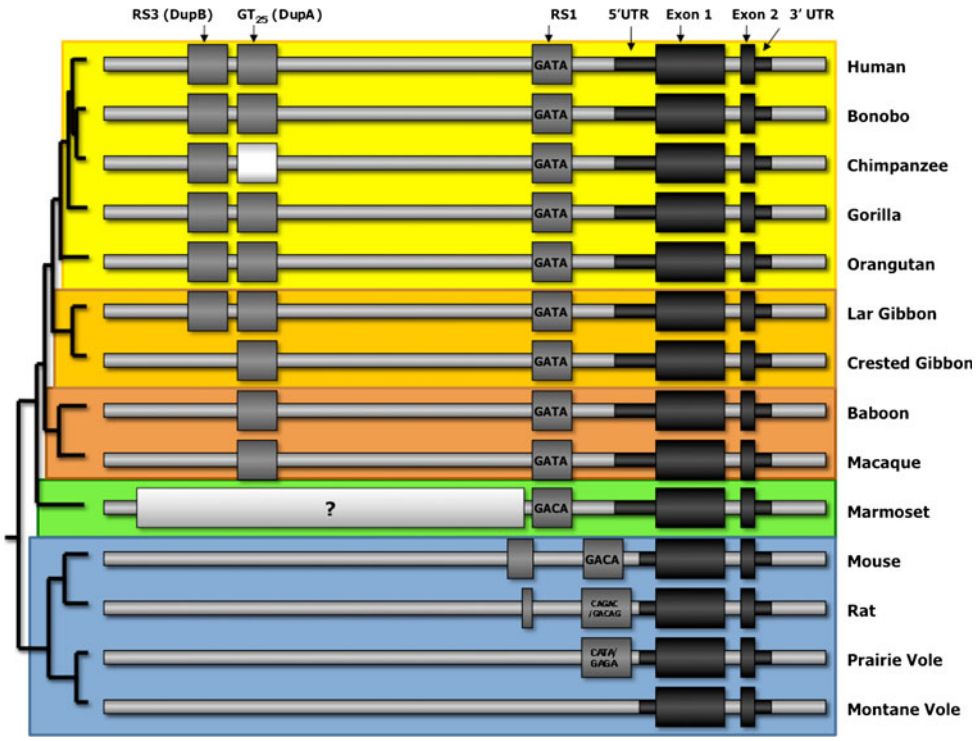


Fig. 1 A phylogenetic representation of the genic structure at the *AVPR1A* locus in different mammalian species. Monogamous voles are known to possess a singular repeat motif known as Repeat Sequence 1 (RS1) of CATA/GAGA that is located roughly 600 nucleotides upstream of the TSS (Hammock and Young 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. Meanwhile, the 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 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 Thibonnier et al. 1996; Hammock and Young 2002, 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 *shaded backgrounds* corresponding to their phylogenetic placement. *Yellow* large bodied apes, *gold* small bodied apes, *orange* Old World monkeys, *green* New World monkeys, *blue* rodents (Color figure online)

(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 24 *A. azarai* individuals for the analysis of intraspecific variation of *AVPR1A* regulatory and coding sequences (Table 1). 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. 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.

We isolated DNA from blood and tissue samples using QIAamp purification columns (Qiagen). All extracted samples were assayed on the NanoDrop ND-1000 spectrophotometer (Thermo Scientific) for quality and quantity, and the DNAs diluted with ddH_2O to a working concentration of 5 ng/μl. The *Aotus* individuals used in this study represented the most disparate mtDNA (maternal) haplotypes in the wild study population (Babb et al., unpublished data).

In addition, we analyzed samples from several other platyrrhine taxa. These included two individuals each from titi (*Callicebus donacophilus*), saki (*Pithecia pithecia*), and squirrel (*Saimiri sciureus*) monkeys (Table 1). 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 Rendell 1994; Boinski

Table 1 Samples investigated at the *AVPR1A* locus

ID	Species	Common name	Sex	Locale
AA008	*A. azarai*	Azara's owl monkey	M	Core area, Formosa, AR
AA014	*A. azarai*	Azara's owl monkey	M	Core area, Formosa, AR
AA015	*A. azarai*	Azara's owl monkey	F	Core area, Formosa, AR
AA021	*A. azarai*	Azara's owl monkey	M	Core area, Formosa, AR
AA032	*A. azarai*	Azara's owl monkey	F	Core area, Formosa, AR
AA034	*A. azarai*	Azara's owl monkey	F	Core area, Formosa, AR
AA037	*A. azarai*	Azara's owl monkey	F	Core area, Formosa, AR
AA053	*A. azarai*	Azara's owl monkey	M	Core area, Formosa, AR
AA057	*A. azarai*	Azara's owl monkey	M	Core area, Formosa, AR
AA063	*A. azarai*	Azara's owl monkey	M	Core area, Formosa, AR
AA067	*A. azarai*	Azara's owl monkey	F	Core area, Formosa, AR
AA071	*A. azarai*	Azara's owl monkey	M	Downstream, Formosa, AR
AA082	*A. azarai*	Azara's owl monkey	M	Downstream, Formosa, AR
AA087	*A. azarai*	Azara's owl monkey	F	Core area, Formosa, AR
AA092	*A. azarai*	Azara's owl monkey	M	Core area, Formosa, AR
AA108	*A. azarai*	Azara's owl monkey	F	Upstream, Formosa, AR
AA109	*A. azarai*	Azara's owl monkey	M	Core area, Formosa, AR
AA114	*A. azarai*	Azara's owl monkey	F	Core area, Formosa, AR
AA122	*A. azarai*	Azara's owl monkey	F	Core area, Formosa, AR
AA123	*A. azarai*	Azara's owl monkey	M	Core area, Formosa, AR
AAF1	*A. azarai*	Azara's owl monkey	F	Saenz-Pena Zoo, AR
AAF1B	*A. azarai*	Azara's owl monkey	F	Saenz-Pena Zoo, AR
AAF2	*A. azarai*	Azara's owl monkey	F	Saenz-Pena Zoo, AR
AAM2	*A. azarai*	Azara's owl monkey	M	Saenz-Pena Zoo, AR
CD01	*C. donacophilus*	White-eared titi monkey	M	San Diego Zoo/CRES
CD02	*C. donacophilus*	White-eared titi monkey	M	San Diego Zoo/CRES
PP01	*P. pithecia*	White-faced saki monkey	P	San Diego Zoo/CRES
PP02	*P. pithecia*	White-faced saki monkey	P	San Diego Zoo/CRES
SS01	*S. sciureus*	Common squirrel monkey	M	San Diego Zoo/CRES
SS02	*S. sciureus*	Common squirrel monkey	F	San Diego Zoo/CRES

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

1999; Rendell and Di Fiore 2007). The DNAs from these species were generously provided by Dr. Oliver Ryder from the Zoological Society of San Diego.

We also conducted searches of the NCBI GenBank database and the University of California-Santa Cruz Genome Browser (UCSC GB) for fully annotated primate and mammalian *AVPR1A* sequences (complete CDS) for use as taxonomic outgroups (Kent et al. 2002; Karolchik et al. 2008). These searches yielded 15 *AVPR1A* protein coding sequences representing 13 species (Table 2). In addition, we retrieved more than 40 *AVPR1A* regulatory/5′ microsatellite sequences from an additional 20 primate taxa for STR motif comparisons.

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) (Fig. 2).

We generated an inter-species *AVPR1A* contig from the sequence files retrieved from GenBank and the UCSC GB using Sequencher, v4.9 (Gene Codes). From the consensus sequence of this assembly, we designed four pairs of *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 (SourceForge.net). These primers amplified four overlapping fragments of ~2000 bp in length, with alternate primers situated within amplicons to facilitate sequencing (Table 3).

Table 2 Comparative samples used in *AVPR1A* sequence analyses

ID	Species	Common name	Chromosomal location of *AVPR1A*	Accession number/ genome assembly
BT01	*Bos taurus*	Cow	5	bosTau4[a]
CF01	*Canis lupus familiaris*	Dog	10	canFam2[a]
CJ01	*C. jacchus*	Marmoset	10 (partial)	calJac1[a]
HS01	*H. sapiens*	Human	12	hg18[a]
MC01	*Macaca mulatta*	Rhesus Macaque	11	rheMac2[a]
MiM01	*Microtus montanus*	Montane Vole	?	EU175984.1[b]
MiO01	*M. ochrogaster*	Prairie Vole	?	EU175983.1[b]
MiP01	*Microtus pinetorum*	Woodland Vole	?	EU175985.1[b]
MiP02	*Microtus pinetorum*	Woodland Vole	?	EU175986.1[b]
MiS01	*Microtus socialis*	Social Vole	?	EU175981.1[b]
MiS02	*Microtus socialis*	Social Vole	?	EU175982.1[b]
MM01	*Mus musculus*	Mouse	10	mm9[a]
PA01	*Pongo albeii*	Orangutan	12	ponAbe2[a]
PT01	*Pan troglodytes*	Chimpanzee	12	panTro2[a]
RN01	*Rattus norvegicus*	Rat	7	rn4[a]

[a] UCSC Genome Browser (Kent et al. 2002; Karolchik et al. 2008)

[b] Fink et al. 2007

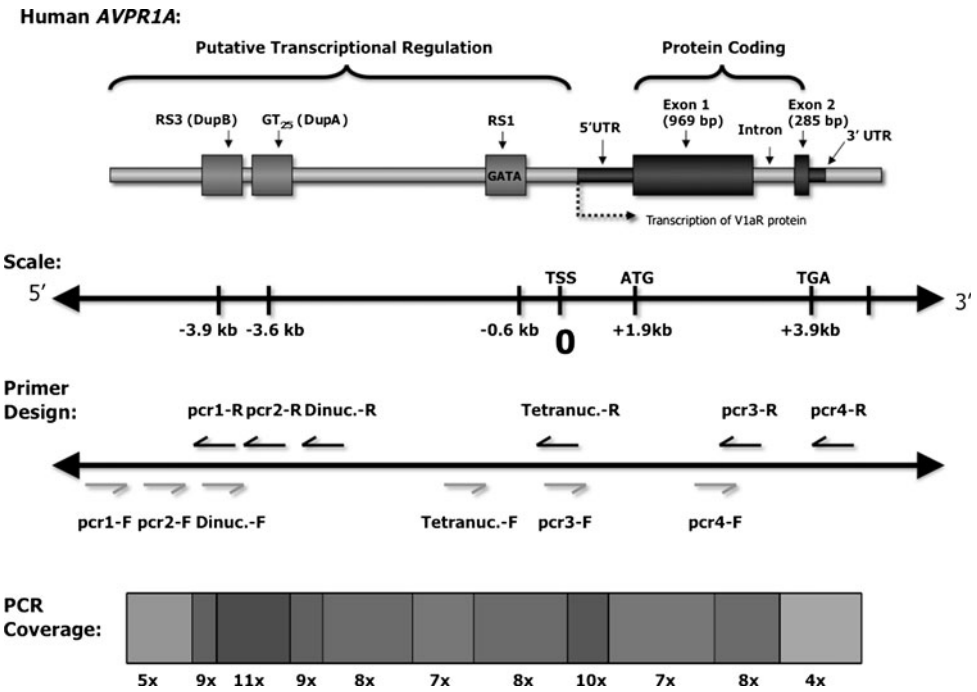


Fig. 2 A schematic diagram of the *AVPR1A* gene in humans. Notable architectural landmarks of this gene are depicted on a scale beneath the figure (Thibonnier et al. 1996). The locations of the oligonucleotide primers used in this study for PCR amplification are shown in their approximate positions given the diagrammatic scale, as well as the coverage achieved by using different primer pairings for DNA amplification, sequencing, and fragment analysis

All owl monkey *AVPR1A* sequence templates were PCR amplified in ABI 9700 Thermocyclers (Applied Biosystems), and visualized through agarose (1% TBE) gel electrophoresis. The amplicons were purified through SAP/Exo I digestion (New England BioLabs), and then cycle sequenced using Big Dye™ v3.1 (Applied Biosystems). Excess dye terminators were removed with the BigDye X-Terminator™ purification kit (Applied Biosystems), and DNA sequences were read on a 3130xl Gene Analyzer (Applied Biosystems). We assessed read quality for each sequence using ABI Sequencing Analysis v5.4 (Applied Biosystems) and aligned them using Sequencher v4.9 (Gene Codes). This approach generated >10.5 kb of reliable *AVPR1A* genic sequence for each individual.

Table 3 Primer sequences for *AVPR1A* amplification and sequencing

Genetic region	Primer name	Strand	Oligonucleotide sequence (5′–3′)
5′ Upstream region (RS3)	pcr1-F[a]	Forward[e]	AGG CAC AGT GGC TCA TAC CT
	pcr1-R[a]	Reverse	GCA AAA CTG CTG ACC ATG AA
5′ Upstream region (RS2)	pcr2-F[a]	Forward[e]	AAC CAT TTA AGT CCC TTC C
	pcr2-R[a]	Reverse	GGT TTT TGG GTA TGC ATT GTG
Coding region (5′UTR/exon 1/intron)	pcr3-F[a]	Forward[e]	AGG ACA AAC ACC GAC GTA GG
	pcr3-R[a]	Reverse	GCC TAC GTG ACC TGG ATG AC
Coding region (exon 1/intron/exon 2)	pcr4-F[a]	Forward[e]	CCG CAG TAC TTC GTC TTC TC
	pcr4-R[a]	Reverse	TCT TCC AAG TCC ATC AAA TTC A
Dinucleotide region (RS2/3)	Dinucleotide-F[b,c]	Forward[e]	GTA TTG CCA CAA ATA GAC CAA CG
	Dinucleotide-R[b,c]	Reverse	GTA AGG ATG ACA GGC GTT ACT G
Tetranucleotide region (RS1)	Tetranucleotide-F[d]	Forward[e]	TAA TAC GAC TCA CTA TAG GG
	Tetranucleotide-R[d]	Reverse	CGC AAG CTT GGC ACT GCG TGC AGC TCT GCT CTG C

[a] This study

[b] Thibonnier et al. (2000)

[c] Hammock and Young (2005)

[d] Kim et al. (2002)

[e] 6-FAM fluorescent versions of each forward primer were also utilized in fragment analysis of *AVPR1A* microsatellite regions

Using the methods described above, we also generated *AVPR1A* genic sequences from two titi monkeys (*C. donacophilus*), two saki monkeys (*P. pithecia*), and two squirrel monkeys (*S. sciureus*). These sequences were used for phylogenetic (interspecific) comparisons of *AVPR1A* variation in primate species.

Sequence Alignments and Repeat Motif Identification

For each individual, we assembled full *AVPR1A* genic sequences by aligning the overlapping forward and reverse fragments using strict (>90%) agreement thresholds. To investigate the different types of variation that could be present at the *AVPR1A* locus, we constructed separate coding region and 5′ regulatory sequence matrices to examine the nucleotide diversity in both the coding and non-coding regions of the gene for all individuals. In total, we obtained >1500 bp of *AVPR1A* coding region (inclusive of exons 1 and 2) and >10500 bp of regulatory sequences from the *A. azarai* individuals.

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., 1F-2R), touchdown 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.

To confirm the sizes of different alleles, putative STR bands were initially observed through gel electrophoresis and then subjected to capillary-based fragment analyses. The fragment analyses utilized ABI AmpliTaq Gold for PCR amplification with a separate set of 6-FAM fluorescent forward primers and the original non-fluorescent reverse oligonucleotide primers used in PCR reactions (Sigma Genosys) (Table 3). Amplicons were visualized through gel electrophoresis using high-resolution 3% NuSieve agarose (Lonza). Allelic sizes were visually and digitally assessed using the CCD-paired Kodak Molecular Imaging (MI) Software v4.5.1 on the Kodak Gel Logic 200 Imaging Station (Carestream Health). Aliquots (1–2 μl) of each PCR amplicon were subsequently combined with 0.3 μl LIZ500 allelic marker standard (Applied Biosystems) and HiDi formamide (Applied Biosystems) to make a 10-μl final reaction volume. Reactions were denatured at 95°C for 3 min and then placed on ice for 3 min. All reactions were then run on a 3130xl Gene Analyzer (Applied Biosystems) for fragment analysis, and alleles were read using GeneMapper ID v4.1 software (Applied Biosystems).

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.

Population Genetic Analysis

Summary statistics, including gene (π) and haplotype (h) diversity, were calculated for *AVPR1A* coding sequences using programs available in Arlequin v3.11 (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 were weighted higher (10) than transitions (1) 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* coding sequences from platyrrhine and other primate species, we restricted the number of *A. azarai* sequences used in the analyses to the most frequent haplotype identified in our study population. This decision was made to avoid biasing the range of sequence variation toward the species *A. azarai*.

We aligned the *A. azarai* coding sequence with the 15 mammalian *AVPR1A* coding sequences obtained from GenBank and UCSC GB, as well as with the sequences from *Callicebus*, *Pithecia*, and *Saimiri*. The resulting matrix of 22 sequences (17 species) was pruned to 563 bases to avoid the ambiguous characters (*N*) found within many of the GenBank sequences. Of the 563 characters in the *AVPR1A* coding region that could be reliably read in all 22 sequences, 374 characters were constant, 55 were phylogenetically uninformative, and 134 were phylogenetically informative.

The resulting *AVPR1A* matrices were 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* 4.0b10 (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.

Network Analysis

We generated multistate median joining (MJ) and binary reduced median (RM) networks from all *A. 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). Here, we used 987 sites from the coding region for this intraspecific analysis, with 10 of them being polymorphic in the 24 individuals. We again employed a 1:10 transition-to-transversion (TI: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 MJ and RM networks for the coding region sequences of *AVPR1A* from comparative primate and mammalian taxa to explore the degree of nucleotide variation in the gene across broader evolutionary distances. In this case, we used the 563 bp sequences available for all species.

Phylogenetic Model Selection

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). We implemented the modified Akaike Information Criterion (AICc) setting due to the small size of comparative nucleotide characters (563 bp), and conducted parallel searches using Bayesian Information Criterion (BIC) and performance-based decision theory (DT). The base tree for our likelihood calculations was optimized for maximum likelihood (ML) phylogenetic analysis.

ML and Bayesian Inference Analyses

We conducted ML analysis in PAUP* 4.0b10 (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 Bayesian inference analysis (BI) with the software program MrBayes version 3.1.2 (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 = 6) with the specific nucleotide substitution model as suggested by jModelTest. The Markov Chain Monte Carlo (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 (Altekar et al. 2004).

We also employed maximum parsimony (MP) and minimum evolutionary distance with neighbor-joining (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.

Analysis of Adaptive Evolution

To identify the level of *AVPR1A* codon variation in our samples, we examined the synonymous-to-non-synonymous amino acid substitution ratio (d_N/d_S) in the first 561 nucleotides of *AVPR1A* exon 1, which represents codons 1–187 in all 22 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.

Bayesian Coalescent Estimation

To check for the presence of *AVPR1A* mutation rate variation among the 17 species, 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 normalized 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 the time points of 31.5 mya (SD 10.0 mya) for the platyrrhine-catarrhine time to most common ancestor (TMRCA), and 20.0 mya (SD 7.5 mya) for the TMRCA of *Aotus*, *Saimiri*, and *Callithrix* (Babb et al., unpublished data).

By implementing the Yule speciation process parameters in BEAST, we specified the abovementioned time points as normally distributed priors applied to the appropriate taxon designations. We used the relaxed-clock log-normal 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 marmoset (calJac1) genome, Due to these architectural gaps, we subsequently broadened our comparison to encompass additional genomes, using the *Homo sapiens* (hg18) 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.

Results

AVPR1A Sequence Diversity in *Aotus*

Our analysis of 24 owl monkeys shows a notable level of diversity in the 1556 bp coding region of the *AVPR1A* gene. A total of 31 polymorphic sites (TIs, TVs, and insertion–deletion substitutions) defining eight distinct haplotypes were present in the coding region (Table 4). The TI:TV ratio was relatively high (13:19 for 32 segregating sites), but both types of mutations were, as expected, proportionally more frequent at the third position. Along the same lines, Nei's gene diversity estimate (π) was 0.004, and the haplotype diversity value was 0.561 among all *A. azarai* individuals. These findings coincide with levels of intraspecific sequence diversity observed in other studies of

Table 4 Summary statistics for *AVPR1A* sequences in *A. azarai*

	A. azarai	
	Coding region	5′ Upstream
Summary statistics		
# Individuals (*n*)	24	24
Nucleotides (bp)	1556	10862
Informative sites	1549	9801
Polymorphic sites	31	53
Transitions (ts)	13	36
Transversions (tv)	19	17
Segregating sites (S)	32	53
Insertion/deletions (indels)	4	1
Haplotypic diversity		
# Haplotypes	8	18
# Singletons	6	15
Haplotype diversity (*h*)	0.562	0.957
Nucleotide diversity		
Nei's gene diversity (π)	0.0435	0.0038
Sequence diversity		
Mean # pairwise differences	67.4022	37.5833
Selective neutrality		
Tajima's *D* (1000 simul.)	27.702	6.5791
P (*D* simul < *D* obs)	0.783	0.002
Fu's F_S (1000 simul.)	−3.6624	−6.1619
P (sim_F_S ≤ obs_F_S)	0.001	0

the *AVPR1A* coding region in mammalian taxa (Fink et al. 2006, 2007; Phelps et al. 2009; Turner et al. 2010).

To investigate the selective neutrality of the *AVPR1A* locus in *A. azarai*, we calculated Fu's F_S and Tajima's *D* statistics for the 24 sequences. This analysis revealed a Tajima's *D* value of 27.702 ($P > 0.7$) and Fu's F_S value of −3.66 ($P < 0.001$) (Table 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 within *A. azarai*. The negative value of Fu's F_S generally confirmed this interpretation.

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 (Fig. 3). *A. azarai* individuals possessed a GACA repeat pattern similar to that of the marmoset at the RS1 location (Fig. 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 4,000 nucleotides upstream of the *A. azarai AVPR1A* start site (ATG). Unfortunately, this region was denoted by 712 "*N*"s in the annotated *C. jacchus* genome sequence (calJac1) 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 GTATATAC repeat in *Callicebus*, *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 *A. azarai* individuals (Fig. 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 (Fig. 1).

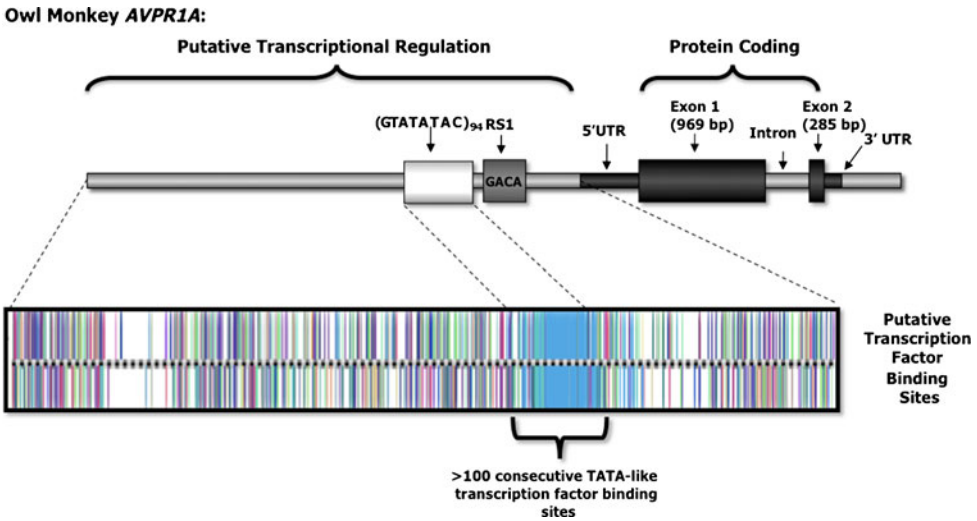


Fig. 3 The structure of the *AVPR1A* locus in *A. azarai*. The localization of putative TFBS is identified by the *boxed area*, which is located upstream of the *AVPR1A* coding sequence. Over 100 consecutive TATA-like transcription factor-binding sites were identified in MatInspector (GenoMatix Software, GmbH)

Table 5 Summary statistics for *AVPR1A* sequences in comparative platyrrhine species

	C. donacophilus		*P. pithecia*		*S. sciureus*	
	Coding region	5′ Upstream	Coding region	5′ Upstream	Coding region	5′ Upstream
Summary statistics						
# Individuals (*n*)	2	2	2	2	2	2
Nucleotides (bp)	563	267	563	267	563	267
Informative sites	563	261	562	266	563	267
Polymorphic sites	1	1	0	1	0	0
Transitions (ts)	0	0	0	1	0	0
Transversions (tv)	1	1	0	0	0	0
Segregating sites (S)	1	1	0	1	0	0
Insertion/deletions (indels)	1	0	0	0	0	0
Selective neutrality						
Tajima's *D* (1000 simul.)	0	0	0	0	0	0
P (*D* simul < *D* obs)	1	1	1	1	1	1
Fu's F_S (1000 simul.)	2.3026	2.3026	0/1	0	0/1	0/1
P (sim_F_S ≤ obs_F_S)	0.107	0.032	1	0.08	1	1

All statistics are based on the number of variants in a given analytical set

AVPR1A Sequence Diversity Among Platyrrhines

The *AVPR1A* protein coding sequences from *Callicebus*, *Pithecia*, and *Saimiri* genera were essentially identical relative to the other individual sequenced from the same species (Table 5). However, each platyrrhine species possessed lineage-specific nucleotide changes that distinguished it from the other genera. In addition, 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 central haplotype, which occurred at high frequencies in the population, and six unique derivative haplotypes extending from it (Fig. 4a). This star-like pattern is indicative of a population expansion and subsequent diversification (Bandelt et al. 1995, 1999; Forster et al. 1996).

We also created networks for the comparison of *AVPR1A* coding sequences from mammalian and primate species. In these networks, the *Aotus* sequence was separated from *Callicebus*, *Pithecia*, and *Saimiri* (Fig. 4b). Among platyrrhine taxa, *Callicebus* and *Pithecia* were clearly distinguished from each other, and both were quite distant from *Saimiri*, whereas *Aotus* showed genetic affinities with *Callithrix*. 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 than the platyrrhines.

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, with respective likelihood scores (−ln *L*) of 2399.09, 2401.31, and 2401.38. Although AICc and BIC selection processes suggested the use of the HKY model plus invariant sites only (+I), the DT search suggested the additional consideration of gamma (+G) distribution for use in our phylogenetic calculations.

ML and Bayesian Inference Analyses

Using the ML phylogenetic algorithm available in *PAUP, we investigated taxonomic relationships of 17 mammalian species based on nucleotide variation in the *AVPR1A* coding sequence. In agreement with the network analysis, the ML tree (Fig. 5a) exhibited a similar breakdown of known taxonomic relationships of species based on genetic

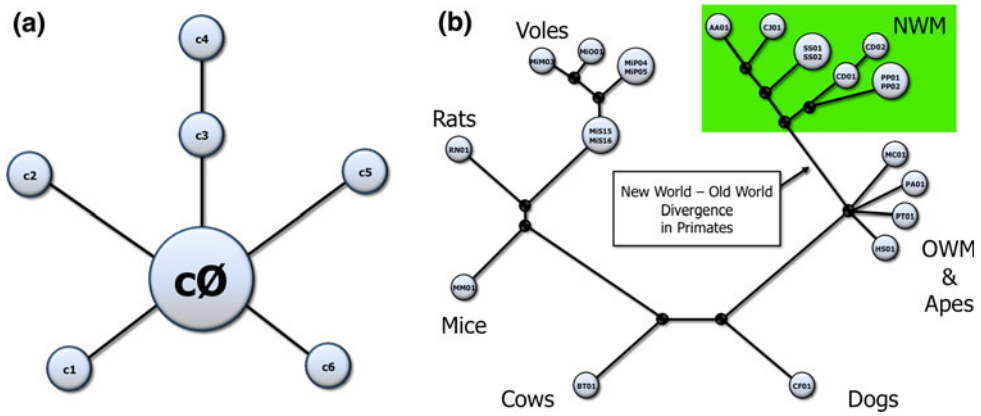


Fig. 4 **a** A MJ network of *AVPR1A* coding region sequences from *A. azarai* individuals. Node size is relative to the number of individuals that share a particular sequence. **b** A MJ network of *AVPR1A* coding region sequences from 17 mammalian and primate species. The branches leading to the different 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

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*, 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).

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) (Fig. 5b). 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.

Bayesian Coalescent Estimation

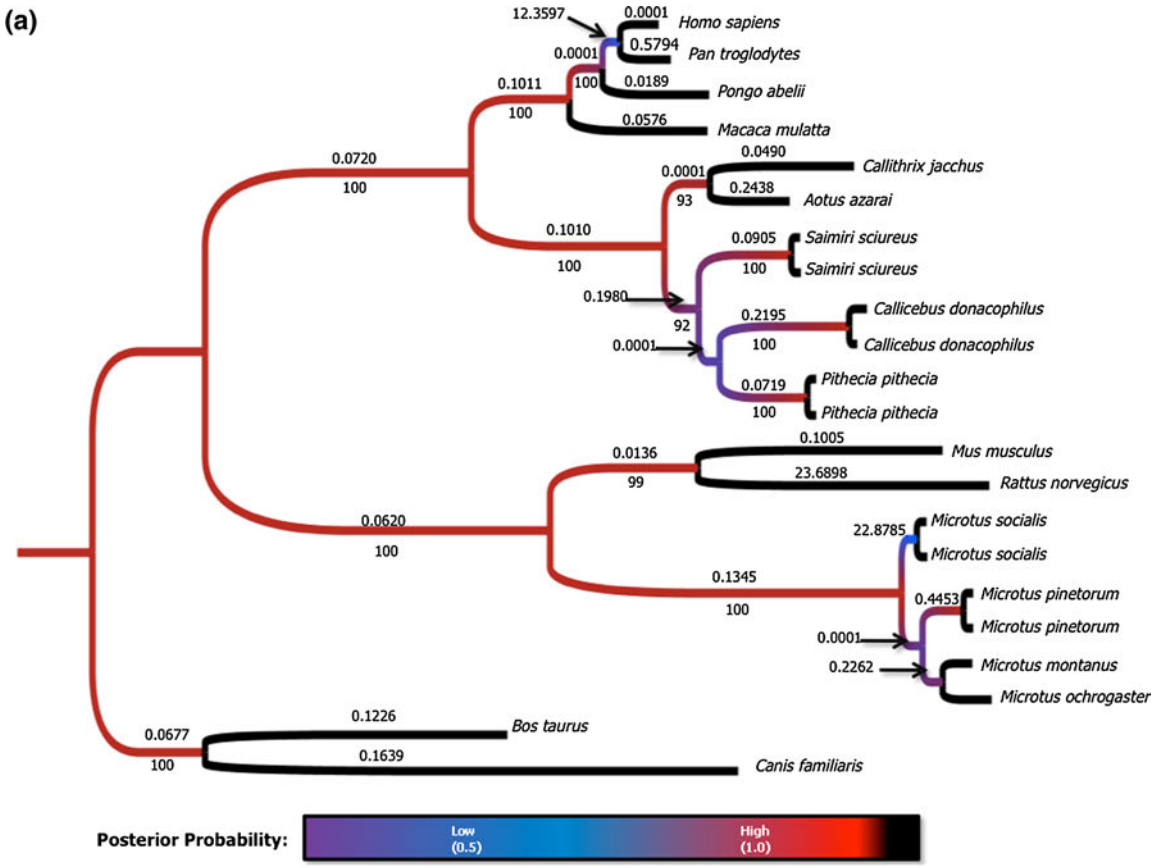
The BEAST analysis produced a highly resolved phylogenetic chronogram of primate and mammalian species with *AVPR1A* coalescent time points for each branch of the tree. Its topology presented a taxonomic arrangement of these species that was very similar to that produced by both the ML and median network analyses (Fig. 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.

Discussion

AVPR1A Evolution in *A. azarai*

The individuals from our owl monkey population exhibit considerable coding region diversity for 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. Thus, there appears to be a lineage specific pattern of sequence diversity at this locus in this species.

One scenario that would explain these data is that varying levels 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 that acting upon the coding region could either conserve or generate novel protein structures. The coding region mutations observed in


(a)
12.3597
0.0001
Homo sapiens
0.5794
Pan troglodytes
0.0001
0.1011
100
100
0.0189
Pongo abelii
0.0576
Macaca mulatta
0.0720
100
0.0490
Callithrix jacchus
0.0001
0.2438
Aotus azarai
93
0.1010
100
0.0905
Saimiri sciureus
100
Saimiri sciureus
0.1980
92
0.2195
Callicebus donacophilus
100
0.0001
Callicebus donacophilus
0.0719
Pithecia pithecia
100
Pithecia pithecia
0.1005
0.0136
Mus musculus
99
23.6898
Rattus norvegicus
0.0620
100
22.8785
Microtus socialis
Microtus socialis
0.1345
100
0.4453
Microtus pinetorum
Microtus pinetorum
0.0001
0.2262
Microtus montanus
Microtus ochrogaster
0.1226
0.0677
Bos taurus
100
0.1639
Canis familiaris
Posterior Probability:
Low (0.5)
High (1.0)
(b)
dN
dS
Homo sapiens
Pan troglodytes
Pongo abelii
Macaca mulatta
Callithrix jacchus
Aotus azarai
Saimiri sciureus
Saimiri sciureus
Callicebus donacophilus
Callicebus donacophilus
Pithecia pithecia
Pithecia pithecia
Mus musculus
Rattus norvegicus
Microtus socialis
Microtus pinetorum
Microtus pinetorum
Microtus montanus
Microtus ochrogaster
Microtus socialis
Bos taurus
Canis familiaris

◀**Fig. 5** **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. **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

Aotus (Fig. 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 (Fig. 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.

In primates, we see a wide range of variation in the *AVPR1A* coding region, and many species exhibit allelic variation outside of the gene in the 5′ upstream region (Fig. 8). However, the repetitive GACA and GTATATAC sequences situated in the 5′ region are invariant in our *A. azarai* population. This finding could imply that selection has acted on the regulatory and coding regions of this gene in different ways in owl monkeys. Alternatively, the unique evolutionary history of the owl monkey genome may have shaped the pattern of molecular diversity that we observe.

Although direct transcriptional assays would be preferable, 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 (Fig. 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

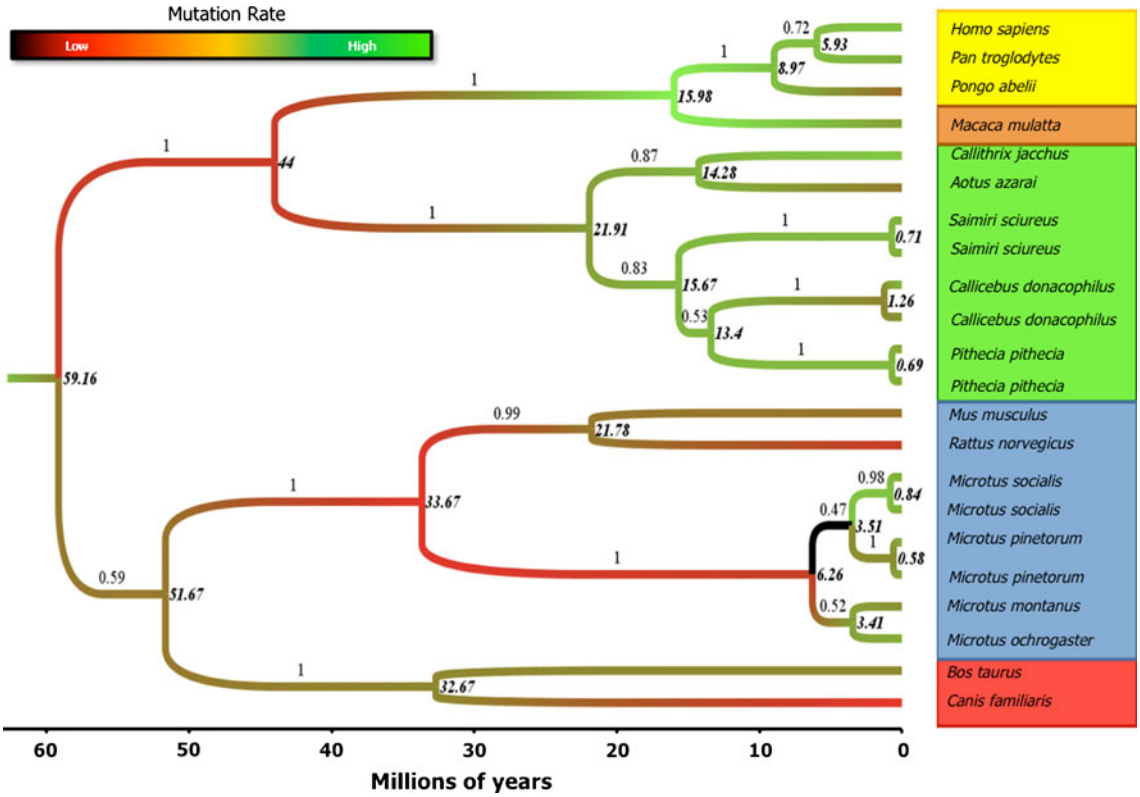


Fig. 6 A chronogram depicting the Bayesian coalescence age estimation of *AVPR1A* coding region sequences, as calculated in BEAST v1.5.3. Age estimates for each node are *italicized* and positioned to the right of them. Posterior probabilities are indicated above each branch. Mutation rate of *AVPR1A* along the different evolutionary lineages is displayed as a *red–green* color gradient across taxa. The different taxa compared here are displayed on *shaded 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) (Color figure online)

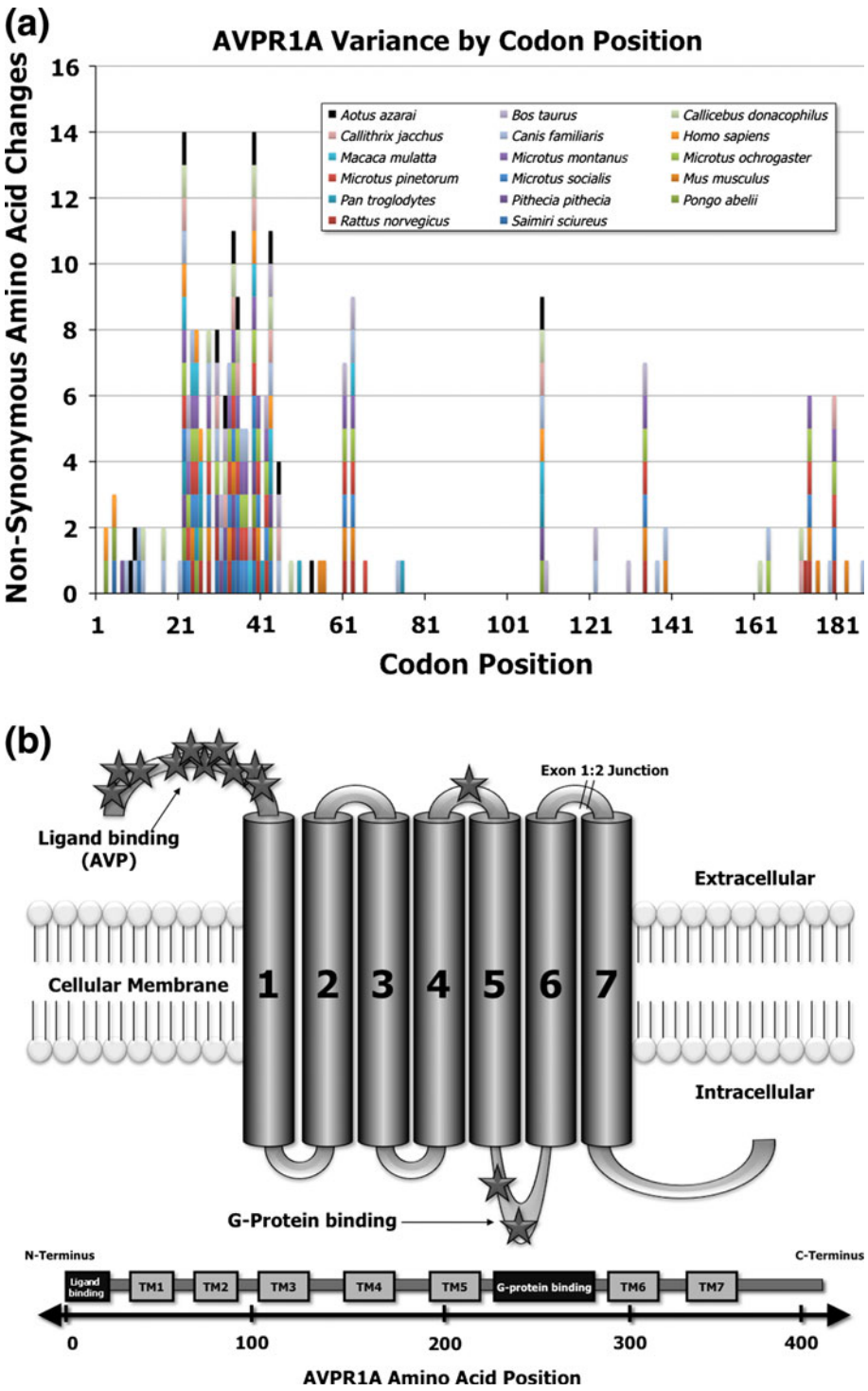


Fig. 7 **a** A stacked histogram displaying the number of non-synonymous amino acid changes of the *AVPR1A* coding sequence in 17 different species, relative to a translated consensus DNA sequence (assessed by plurality, [50% majority rule]). *Black bars* represent the location of these changes in *A. azarai*, whereas grey bars represent transcriptional changes in the other taxa examined. 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). **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 (**a**) and simulated in Topo2 (adapted from Fink et al. 2007) (Color figure online)

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

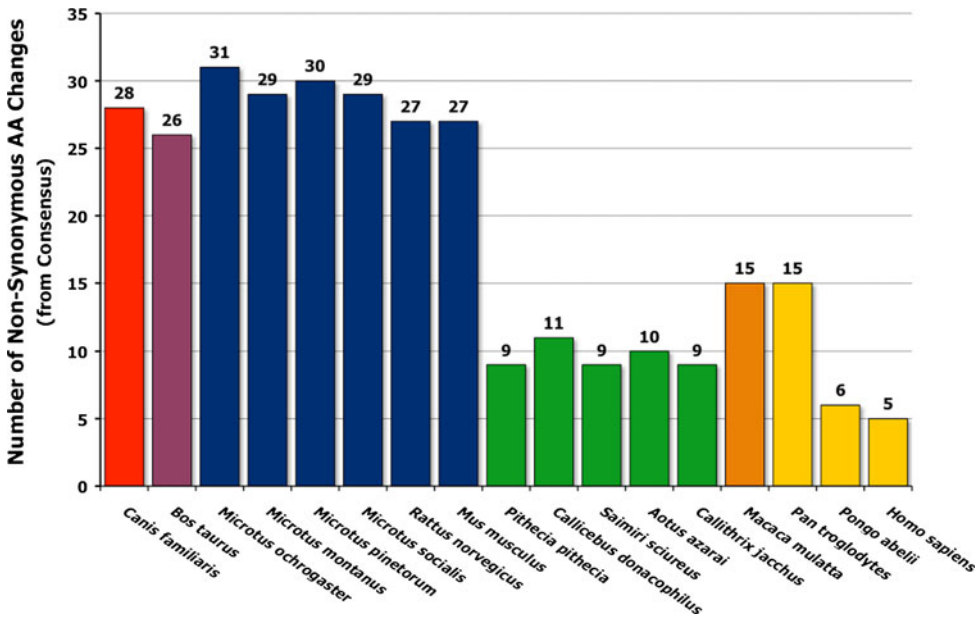


Fig. 8 The frequency of non-synonymous amino acid changes in the *AVPR1A* gene, as calculated from the consensus translation (by plurality) of 17 mammalian and primate sequences

experience different 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 (Fink et al. 2007) with regard to the N-terminus of the mature V1a receptor protein, 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 five platyrrhine genera examined in this study (*Aotus*, *Callicebus*, *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 (Figs. 4b, 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; Rendell and Di Fiore 2007).

That selection may be at work at this locus is indicated by the statistical analysis of *AVPR1A* coding sequences in *A. azarai*. 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. However, these same values could reflect a demographic subdivision in a population's recent history, rather than selection itself. The latter hypothesis is refuted by our analysis of mtDNA diversity of this same population, which indicates that it has remained constant in size over time (Babb et al., unpublished data). Thus, selection appears to have shaped diversity at this locus in *A. azarai*, although it is unclear whether the coding region is the specific target of selection.

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 mammalian 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 (hg18, chr12: 61826483–61832857) (Fig. 9). In addition, when surveying this region using the DGV (Iafrate et al. 2004) track on UCSC GB, we noted that humans, chimps, and rhesus macaques display copy number variants (CNVs) in large sections of the entire region. Thus, a large degree of transcriptional plasticity and differential gene dosages may occur at this locus in primate species.

Along the same lines, the *AVPR1A* 5′ regulatory region has been shown to be highly variable, even in closely

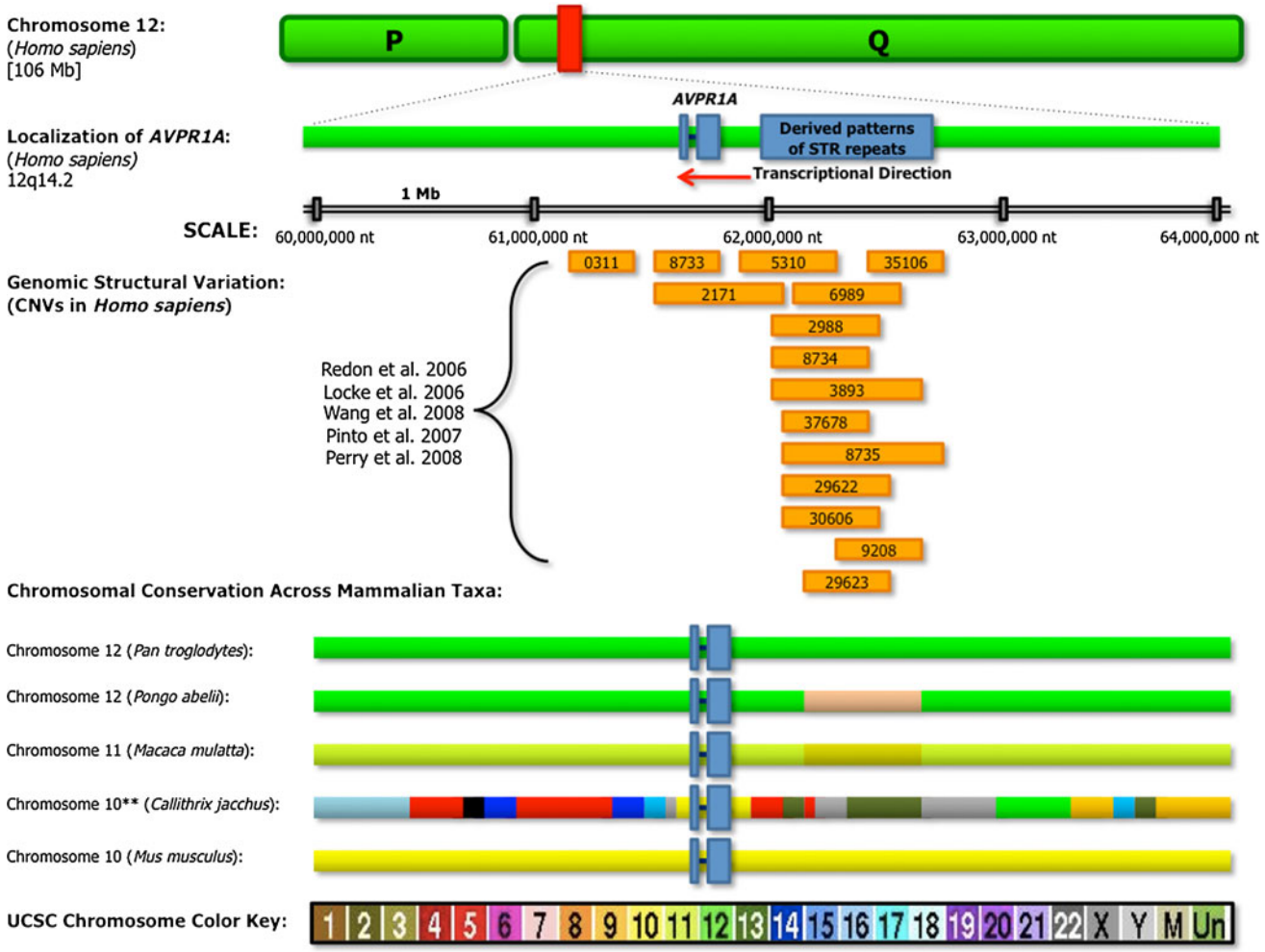


Fig. 9 Genomic structural variation near the *AVPR1A* gene in different mammalian species. The particular chromosomal location of the gene has changed multiple times throughout mammalian evolution, as shown by the chromosome maps on the lower portion of the figure (adapted from UCSC GB [Kent et al. 2002; Karolchik et al. 2008])

related taxa (Kim et al. 2002; Hammock and Young 2005; Donaldson et al. 2008; Rosso et al. 2008). Our data for 5′ *AVPR1A* microsatellite variation in *Aotus*, *Saimiri*, *Callicebus*, and *Pithecia* clearly support these findings. Since this 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, this putative regulatory sequence may, in fact, be the molecular element under selective pressure in mammalian taxa, rather than simply the coding region itself.

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 (by the way of 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.

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 in cerocopithecine and hominoid species. These data provide new clues into 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.

Acknowledgments The authors would like to thank Dr. Ömer Gökçümen for his assistance with statistical calculations, Dr. Richard Smith for advice with transcription factor mapping, and two anonymous reviewers whose comments on earlier versions of this manuscript were invaluable. This study was supported by a University Research Fund award from the University of Pennsylvania (EFD), and research funds from the Department of Anthropology at the University of Pennsylvania (PLB). E.F.D. also acknowledges continuing financial support from the Wenner-Gren Foundation, the L.S.B. Leakey Foundation, the National Geographic Society, the National Science Foundation (BCS-0621020), and the Zoological Society of San Diego. T.G.S. further acknowledges the infrastructural support provided by the National Geographic Society.

References

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

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, 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, 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

Bell SD, Magill CP, Jackson SP (2001) Basal and regulated transcription in Archaea. Biochem Soc Trans 29(4):392–395

Boinski S (1999) The social organizations of squirrel monkeys: implications for ecological models of social evolution. Evol Anthropol 8(3):101–114

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(7289):704–712

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

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) Touchdown PCR to circumvent spurious priming during gene amplification. Nucl Acids Res 19(14):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(8):180

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

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

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinfor 1:e47–e50

Felsenstein J (2005) PHYLIP (phylogeny inference package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle (USA). http://evolution.genetics.washington.edu/phylip.html

Fernandez-Duque E (2009) Natal dispersal in monogamous owl monkeys (*Aotus azarai*) of the Argentinean Chaco. Behaviour 146:583–606

Fernandez-Duque E, Rotundo M (2003) Field methods for capturing and marking Araza's night monkeys. Int J Primat 24:1113–1120

FigTrees v1.2.3 (2009) (A. Rambaut, http://beast.bio.ed.ac.uk/FigTree)

Fink S, Excoffier L, Heckel G (2006) Mammalian monogamy is not controlled by a single gene. Proc Natl Acad Sci USA 103(29):10956–10960

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

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

Gökçümen Ö, Lee C (2009) Copy number variants (CNVs) in primate species using array-based comparative genomic hybridization. Methods 49(1):18–25

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

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

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 (2005) Microsatellite instability generates diversity in brain and sociobehavioral traits. Science 308(5728):1630–1634

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(5):289–301

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

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

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17(8):754–755

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294(5550):2310–2314

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(9):949–951

Insel TR (2003) Is social attachment an addictive disorder? Physiol Behav 79(3):351–357

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. Nucl Acids Res 36:773–779

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at UCSC. Genome Res 12(6):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(5):503–507

Kleiman DG (1977) Monogamy in mammals. Q Rev Biol 52:39–69

Korbie DJ, Mattick JS (2008) Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nat Protoc 3(9):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

Lim MM, Hammock EA, Young LJ (2004a) The role of vasopressin in the genetic and neural regulation of monogamy. Neuroendocrin 16(4):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(6993):754–757

Maddison DR, Maddison WP (2003) MacClade 4: analysis of phylogeny and character evolution, Version 4.06. Sinauer and Associates, Sunderland, MA

Mendoza SP, Reeder DM, Mason WA (2002) Nature of proximate mechanisms underlying primate social systems: simplicity and redundancy. Evol Anthropol 11(S1):112–116

Mock DW, Fujioka M (1990) Monogamy and long-term pair bonding in vertebrates. Trends Ecol Evol 5(2):39–43

Moller AP (2003) The evolution of monogamy: mating relationships, parental care and sexual selection. In: Reichard UH, Boesch C (eds) Monogamy: mating strategies and partnerships in birds, humans and other mammals. University of Cambridge Press, Cambridge, pp 29–41

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(34):20042–20050

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(1):274–280

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

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(5):400–405

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

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(18):7392–7396

Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25(7):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. Syst Biol 55(2):228–244

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

Rendell D, Di Fiore A (2007) Homoplasy, homology, and the perceived special status of behavior in evolution. J Hum Evol 52:504–521

Ronquist F, Huelsenbeck JP (2003) MrBayes 3, Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Rosso L, Keller L, Kaessmann H, Hammond RL (2008) Mating system and *avpr1a* promoter variation in primates. Biol Lett 4(4):375–378

Rotundo M, Fernandez-Duque E, Dixson A (2005) Infant development and parental care in free-ranging *Aotus azarai azarai* in Argentina. Int J Primat 26:1459–1473

Rozas J, Sanchez-Delbarrio JC, Messeguer X, Rozas R (2003) DnaSP: DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Swofford DL (2002) PAUP*, phylogenetic analysis using parsimony (and other methods), Version 4.0b10 Beta. Sinauer Associates, Sunderland, MA

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

Tejedor MF, Tauber AA, Rosenberger AL, Swisher CC, Palacios ME (2006) New primate genus from the Miocene of Argentina. Proc Natl Acad Sci USA 103(14):5437–5441

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

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(6):1269–1278

Van Schaik CP, Dunbar RIM (1990) The evolution of monogamy in large primates: a new hypothesis and some crucial tests. Behaviour 115(1–2):30–61

Van Schaik CP, Kappeler PM (2003) The evolution of social monogamy in primates. In: Reichard UH, Boesch C (eds) Monogamy: mating strategies and partnerships in birds, humans

and other mammals. University of Cambridge Press, Cambridge, pp 59–80

Van Schaik CP, Van Hoof JARAM (1983) On the ultimate causes of primate social systems. Behaviour 85(1–2):91–117

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(37):14153–14156

Wildman DE, Jameson NM, Opazon JC, Yi SV (2009) A fully resolved genus level phylogeny of neotropical primates (Platyrrhini). Mol Phy Evol 53(3):694–702

Wright PC (1994) The behavior and ecology of the owl monkey. In: Baer JF, Weller RE, Kakoma I (eds) Aotus: the owl monkey. Academic Press, San Diego, CA, pp 97–113

Yang Z (2007) PAML 4: a program package for phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Young LJ, Hammock EA (2007) On switches and knobs, microsatellites and monogamy. Trends Genet 23(5):209–212

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(1):51–57