Short poly-glutamine repeat in the androgen receptor in New World monkeys

Chihiro Hiramatsu[a,b,*], **Annika Paukner**[c], **Hika Kuroshima**[d], **Kazuo Fujita**[d], **Stephen J. Suomi**[c], and **Miho Inoue-Murayama**[e,f]

Chihiro Hiramatsu: chihiro@design.kyushu-u.ac.jp; Annika Paukner: pauknera@mail.nih.gov; Hika Kuroshima: kuroshima.hika.4z@kyoto-u.ac.jp; Kazuo Fujita: kfujita@bun.kyoto-u.ac.jp; Stephen J. Suomi: suomis@lce.nichd.nih.gov; Miho Inoue-Murayama: mmurayama@wrc.kyoto-u.ac.jp

[a]Department of Human Science, Faculty of Design, Kyushu University, Postal address: 4-9-1 Shiobaru, Minamiku, Fukuoka, 815-8540, Japan

[b]Physiological Anthropology Research Center, Kyushu University, Postal address: 4-9-1 Shiobaru, Minamiku, Fukuoka, 815-8540, Japan

[c]Laboratory of Comparative Ethology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health Postal address: Elmer School RD Room 217, Poolesville, MD 20837, USA

[d]Department of Psychology, Graduate School of Letters, Kyoto University, Postal address: Yoshidahonmachi, Sakyo-ku, Kyoto 606-8501, Japan

[e]Wildlife Research Center, Kyoto University, Postal address: Tanaka-Sekiden-cho, Sakyo-ku, Kyoto 606-8203, Japan

[f]Wildlife Genome Collaborative Research Group, National Institute for Environmental Studies, Postal address: 16-2, Onogawa, Tsukuba, Ibaraki 305-8506, Japan

Abstract

The androgen receptor mediates various physiological and developmental functions and is highly conserved in mammals. Although great intraspecific length polymorphisms in poly glutamine (poly-Q) and poly glycine (poly-G) regions of the androgen receptor in humans, apes and several Old World monkeys have been reported, little is known about the characteristics of these regions in New World monkeys. In this study, we surveyed 17 species of New World monkeys and found length polymorphisms in these regions in three species (common squirrel monkeys, tufted capuchin monkeys and owl monkeys). We found that the poly-Q region in New World monkeys is relatively shorter than that in catarrhines (humans, apes and Old World monkeys). In addition, we observed that codon usage for poly-G region in New World monkeys is unique among primates. These results suggest that the length of polymorphic regions in androgen receptor genes have evolved uniquely in New World monkeys.

[*]Corresponding author: Chihiro Hiramatsu, Department of Human Science, Faculty of Design, Kyushu University, 4-9-1 Shiobaru, Minamiku, Fukuoka, 815-8540, Japan, chihiro@design.kyushu-u.ac.jp, phone: +81 92-553-4456.

Introduction

The androgen receptor (AR) is a ligand-activated nuclear hormone receptor that plays a critical role in the development of the male phenotype (Gelmann, 2002; Li and Al-Azzawi, 2009). A dimer of the receptors works as a transcription factor in the nucleus when it is combined with internal androgen hormones, testosterone and its metabolite, 5α-dihydrotestosterone. AR is almost ubiquitously expressed in tissues and mediates various physiological and developmental functions in addition to virilization. Androgens in the brain are involved in circuit formation during development and maintaining proper function in adulthood. Thus, expression of AR is considered to affect not only reproductive but also non-reproductive behavior patterns dealing with aggression and emotion (Li and Al-Azzawi, 2009).

The androgen receptor gene (*AR*) is located on the X-chromosome and consists of eight exons in humans (Brown et al., 1989; Kuiper et al., 1989). The similar gene structure in mice, monotremes and marsupials indicates *AR* is highly conserved in mammals (Choong et al., 1998; Faber et al., 1991; He et al., 1990; Migeon et al., 1981; Spencer et al., 1991). The N-terminal domain of AR is coded in exon 1 and relates to transcription regulation by *AR*. Other exons code DNA binding domain and ligand binding domains (Fig. 1) (Li and Al-Azzawi, 2009; Quigley et al., 1995).

In exon 1, there are trinucleotide microsatellite repeat regions, (CAG)n and (GGN)n. These repeat regions yield poly glutamine (poly-Q) and poly glycine (poly-G) repeats in the N-terminal domain. The poly-Q region is further segmented into QI and QII regions (Fig. 1). Significant variation in the length of poly-QI is known in humans. Normal repeat lengths of QI range approximately from 10 to 37 and the most frequent repeat length is 23 (Hsing et al., 2000; Huhtaniemi et al., 2009). Shorter repeats of poly-QI increase the risk of prostate cancer and extremely long repeats (> 40) are related to spinal and bulbar muscular atrophy and a variable degree of androgen insensitivity although causation is controversial among studies (Rajender et al., 2007). The length polymorphism of poly-G region is also found in humans (ranges approximately from 14 to 27 repeats), while no great variation was reported for poly-Q II.

The length polymorphisms in these regions are also found in non-human primates. Interestingly, if a species is phylogenetically closer to humans, it possesses longer repeats and higher polymorphism comparable to humans (Choong et al., 1998; Hong et al., 2006; Mubiru et al., 2012). For example, our closest relative, the chimpanzee, has length polymorphism ranging from 15 to 27 at poly-QI but no polymorphism was found in rhesus macaques (Hong et al., 2006; Mubiru et al., 2012). The tendency of allele combination between longer poly-QI and shorter poly-G repeats in chimpanzees (Hong et al., 2006) and insusceptibility to idiopathic infertility in haplotypes with longer poly-QI and shorter poly-G combination in humans (Ferlin et al., 2004) indicates the underlying constraints on the

combination of these regions. The association between these polymorphisms and personality traits such as aggressiveness has also been an intriguing topic of personality research on non-human animals as well as humans (Jonsson et al., 2001; Konno et al., 2011; Rubinow and Schmidt, 1996).

Little is known, however, about the characteristics of these regions in other primates. Only one species of New World monkeys (platyrrhine), the marmoset, and several species of prosimians have been surveyed so far and no polymorphism has been found in these taxa (Mubiru et al., 2012; Poux et al., 2005). Hormonal studies have reported that the androgen concentration in several species of New World monkeys is higher than that in Old World monkeys (Snipes et al., 1969), and their androgen concentration is associated with social status (Bales et al., 2006; Schoof and Jack, 2013). These findings indicate that androgen plays an important role in behavior of New World monkeys and highlights the importance of understanding the structure and variation of AR in this taxon. Therefore, we aimed to characterize the *AR* gene in New World monkeys and to understand the entire picture of AR variation in primates by focusing on length polymorphism of AR poly-Q and poly-G regions and codon usage bias in 17 species of New World monkeys.

Materials and Methods

DNA samples

DNA samples were obtained from 120 individuals that belong to one of 17 species of New World monkeys. Number of individuals in each species surveyed is summarized in Table 1. Most samples were obtained from animals kept at various sites in Japan (Japan Monkey Center; Primate Research Institute, Kyoto University; Graduate School of Letters, Kyoto University; Kyoto City Zoo and Omuta City Zoo). DNA was extracted from tissues when animals were dead, or from blood or hairs when animals were alive by using QIAamp DNA mini Kit (Qiagen, California, USA). For 22 capuchin monkeys, DNA extracted from blood samples of monkeys kept at National Institute of Child Health and Human Development (NICHD) were used. For individuals whose sex was not known at sampling, we determined their sex by PCR amplification of sex-determining region Y (*SRY*). The 22 tufted capuchins in NICHD and nine tufted capuchins, four common squirrel monkeys and six Bolivian squirrel monkeys at Kyoto University were kept in several groups at each site and some individuals were related to each other. No clear genealogy and kinship were available for other individuals. We have complied with the ethical standards in the treatment of animals with the guidelines laid down by the Primate Society of Japan. The experiment was approved by the Animal Experiments Committee of the Graduate School of Letters (16–37) and the Wildlife Research Center (WRC-2016-002A), Kyoto University, and the Animal Care and Use Committee of NICHD (ASP#12-064) in accordance with the European Directive 2010/63 on the Protection of Animals in Scientific Experimentation.

Determination of repeat length and sequencing

Poly-Q and poly-G regions were PCR amplified by using following primer sets. Poly-Q: ARhf 5'-TCCAGAATCTGTTCCAGAGCGTGC-3' and ARhr 5'-GCTGTGAGGGTTGCTGTTCCTCAT-3'. Poly-G: ARGFH 5'-

CAGTGCCGCTATGGGGACCTGGCGA-3' and ARGR 5-
GGACTGGGATAGGGCACTCTGCTCACC-3'.

To determine the length of repeated region, forward primers were fluorescent labeled (FAM for ARhf and HEX for ARGFH). A PCR was carried out with approximately 20ng genomic DNA, forward and reverse primers 0.5uM each, 0.4mM each for dNTP mix mixture, LA Taq polymerase (Takara bio, Shiga, Japan) 0.5 unit, 1×Takara PCR buffer and ddH2O at 10ul scale. PCR condition was as follows: after initial incubation at 95°C for 2min, 35 cycles of 95°C for 30 sec, 55°C for 30 sec, 74°C for 2min, followed by a final extension at 74°C for 10 min. After the PCR amplification, PCR samples were diluted a hundred times with ddH2O and subjected to electrophoresis by Applied Biosystems 3130xl Genetic Analyzer with Gene Mapper software (Applied Biosystems, Foster City, California, USA). After determination of the repeat length, we chose individuals who were hemi- or homozygous for these repeat sites and the sequence was determined for each species by using Applied Biosystems 3130xl Genetic Analyzer. For the species that had polymorphism at these repeat regions, the sequence was determined for each allele from hemi- or homozygous individuals if available. We aligned sequences of a representative allele (most frequent allele if polymorphic) from each species including previously reported primates and house mouse using CLUSTAL W (Thompson et al., 1994) implemented in MEGA version 5 (Tamura et al., 2011). We counted the number of glutamine and glycine coded in the region corresponding to the poly-QI, poly-QII and poly-G regions of other primates for which sequences were previously reported. We included all glutamine or glycine sites in repeat numbers even when other amino acids were inserted in the corresponding regions.

Calculation of codon bias and visualization of similarity among species

As an indicator of codon bias, we calculated the ratio of each codon for poly-Q and poly-G regions in representative alleles if the representative sequence was available. To visualize the relationship of these regions among species from various taxa, we applied multi-dimensional scaling (MDS) analysis using the nine variables (six codon usage ratios, length of poly-QI, poly-QII and poly-G in the representative alleles). The z-scored variables were used for the analysis to normalize the range of each variable.

Relationship between poly-QI and poly-G lengths

To inspect the possibility that longer poly-QI and shorter poly-G are linked under the constraints of transcriptional functions as suggested for humans and apes (Ferlin et al., 2004; Hong et al., 2006), we examined the relationship between length of poly-QI and poly-G in alleles of New World monkeys. The combinations of poly-QI and poly-G were obtained from individuals who were homozygous on either region. For comparison, we also replicated this analysis with data from great apes (gorillas, orangutans, chimpanzees and bonobos) obtained in Hong et al. 2006. Combinations in each species were shown in three dimensional plots to visualize the frequency of each combination as elevation.

Results

Length, polymorphism and sequence of repeated regions

The alignment of representative sequences from all species is illustrated in Fig. 2. For species that possess polymorphisms, the representative allele indicates the most frequent allele observed in the species. Details of polymorphic sites found in this study are shown in Fig. 2. The sequences of the repeat regions obtained in this study are available at the DDBJ/EMBL/GenBank nucleotide sequence database with the accession numbers from LC185460 to LC185496 and LC186017. Accession numbers of the sequences in Fig. 2 are summarized in Table 1. It appears that poly-QI in New World monkeys tends to be shorter compared to catarrhine species (human, apes and Old World monkeys) (Fig. 2). In New World monkeys, there was no length polymorphism in poly-QI where the extensive length polymorphisms were observed in apes (Choong et al., 1998; Djian et al., 1996; Garai et al., 2014; Hsing et al., 2000; Huhtaniemi et al., 2009; Mubiru et al., 2012). In poly-QII, we found a length polymorphism, 4 or 5 CAG repeats in capuchin monkeys. The repeat length of the poly-G region in New World monkeys tended to be shorter than in catarrhines except for squirrel monkeys that possess long G repeats comparable to humans and apes. We found length polymorphism in squirrel monkeys (21, 22 and 23 repeats), in capuchin monkeys (11, 13 and 14 repeats) and in owl monkeys (11 and 14 repeats) at this region (Table 2). The sequences of 22 and 23 poly-G repeats in squirrel monkeys were not obtained since there were no hemi- or homozygous individuals for these alleles. The sequence of 14 poly-G repeats in capuchin monkeys has not been sequenced due to the sample availability. In addition to length polymorphisms, we also found one nonsynonymous single nucleotide polymorphism that codes serine (codon: AGT) or glycine (codon: GGT), at the 91th amino acid (the 4th amino acid after Poly-Q II) in owl monkeys.

Frequency and heterozygosity of polymorphic alleles

For the species in which length polymorphisms were found, frequencies of each allele are shown in Table 3. The chi-square test for Hardy-Weinberg equilibrium showed that there is no reason to reject the hypothesis of random mating for all populations (all $p > 0.05$). The expected heterozygosity calculated from frequencies, and the observed heterozygosity in females, are also provided in Table 3.

Codon usage

The codon usage of poly-Q and poly-G regions in the representative sequence of each species is shown in Table 4. It is notable that high codon bias for GGA codon in the poly-G region was observed in New World monkeys especially in Callitrichidae species (tamarin and marmoset) although the codon was seldom observed in other taxa. The Euclidean distances between species by MDS analysis were plotted in a two-dimensional space (Fig. 3). New World monkeys were clearly separated from the groups of great apes and prosimians. Within New World monkeys, tamarins were separated from other species, probably reflecting their unique codon bias.

Relationship between poly-QI and poly-G repeat length

We obtained 15 poly-QI and poly-G combinations of common squirrel monkeys, 71 combinations of tufted capuchins and 11 combinations of owl monkeys. We also replicated the analysis for 27 combinations of gorillas, 37 combinations of orangutans, 209 combinations of chimpanzees and 4 combinations of bonobos. Fig. 4 shows the relationship between poly-QI and poly-G repeat lengths in New World monkeys (filled symbols) and in great apes (open symbols). The elevation of each plot (indicated by a dashed line) represents the frequency of each combination in a given species. For species without polymorphism, elevation was set to be 1. Hong and colleagues showed that there was a negative correlation between poly-Q I and poly-G length in chimpanzees (Hong et al., 2006) and our analyses replicate the result ($r = -0.38$, $p < 0.0001$, Pearson correlation), while no such relationship was observed in other species. The plots show clearly less length variation at poly-QI compared to poly-G among New World monkeys. It appears that poly-QI lengths in New World monkeys are shorter than those in apes (Fig. 4).

Discussion

We found polymorphisms in three species and observed that the poly-Q region is short in New Word monkeys. We also found unique codon bias in this taxon. Although great variation of CAG and GGN repeat length in the androgen receptor gene has been reported in many primate species, little was known about the variation in New World monkeys except for common marmosets (synonym: white-tufted-ear marmoset)(Mubiru et al., 2012). In this study, we surveyed 17 species of New World monkeys and found length polymorphisms at these regions in three species (squirrel monkeys, capuchin monkeys, and owl monkeys). Frequencies of each allele and the expected and the observed heterozygosity indicate the existence of large diversity in these polymorphisms. Since the sample size of common squirrel monkey and owl monkey is small, and some individuals are related to each other in capuchins and squirrel monkeys, caution should be taken regarding the interpretation of the extent of diversity. We did not observe polymorphism either in Q repeat nor G repeat regions in Callitrichidae, tamarins and marmosets. This observation is consistent with a previous report that showed no length polymorphism in common marmosets (Mubiru et al., 2012). For species that possess polymorphisms, it would be interesting to examine if there are functional differences among alleles.

Although our sample size is not enough to conclude that there is no great genetic variation of the AR gene in most New World monkeys, it would be reasonable to state that the microsatellite region in the exon 1 of the androgen receptor gene of New World monkeys (platyrrhine) is relatively shorter than that in catarrhines. In addition, we observed that codon usage at the G repeat region in New World monkeys, especially in Callitrichidae, is unique among primates. GGA at the poly-G region is frequently used in tamarins and squirrel monkeys and used once in marmoset, capuchins, spider monkey and howler monkeys while the codon is never used in other taxa except for aye-aye (Fig. 2, Table 4). This suggests that the poly-G region has evolved uniquely in New World monkeys. The MDS analysis that shows the close relationship among New World monkeys supports this idea (Fig. 3). Uniqueness of New World monkeys in polymorphic regions of a functional gene comparing

to Old World primates also has been reported in the serotonin transporter gene (Inoue-Murayama et al., 2008; Santangelo et al., 2016). The difference in androgen concentration between New World monkeys and Old World monkeys (Snipes et al., 1969) might be related to the regulation of androgen via AR. Thus, the relationship between sequence differences and functional differences of the *AR* gene among primates needs to be investigated in future research.

It can be hypothesized that the length of poly-Q and poly-G regions should be balanced since AR works as a dimer, i.e. two identical molecules linked together (Li and Al-Azzawi, 2009). Although a negative correlation between poly-QI and poly-G lengths was observed in chimpanzees (Hong et al., 2006), no such relationship was observed in New World monkeys (Fig. 4). In humans, the combination between long Q and long G repeats seems associated with male infertility (Ferlin et al., 2004). These lines of evidence may suggest that the balancing between poly-Q and poly-G regions is important only when these repeats are too long as in the case of apes and humans.

The nature of longer microsatellites in humans compared to other primates is one of the intriguing evolutionary questions. Although several mechanisms, such as differences in mutation ratio, reproductive age, population size etc. have been hypothesized (Rubinsztein et al., 1995), a clear explanation has not been put forward. The tendency of a short poly-Q region in New World monkeys shown in this study may provide new insight into the evolution of microsatellites in a functional gene in primates.

Acknowledgments

References

Bales KL, French JA, McWilliams J, Lake RA, Dietz JM. Effects of social status, age, and season on androgen and cortisol levels in wild male golden lion tamarins *(Leontopithecus rosalia)*. Horm Behav. 2006; 49:88–95. DOI: 10.1016/j.yhbeh.2005.05.006 [PubMed: 15978593]

Brown CJ, Goss SJ, Lubahn DB, Joseph DR, Wilson EM, French FS, Willard HF. Androgen receptor locus on the human X chromosome: regional localization to Xq11-12 and description of a DNA polymorphism. Am J Hum Genet. 1989; 44:264–269. [PubMed: 2563196]

Choong CS, Kemppainen JA, Wilson EM. Evolution of the primate androgen receptor: A structural basis for disease. J Mol Evol. 1998; 47:334–342. DOI: 10.1007/Pl00006391 [PubMed: 9732460]

Djian P, Hancock JM, Chana HS. Codon repeats in genes associated with human diseases: fewer repeats in the genes of nonhuman primates and nucleotide substitutions concentrated at the sites of reiteration. Proc Natl Acad Sci U S A. 1996; 93:417–421. [PubMed: 8552651]

Faber PW, King A, Vanrooij HCJ, Brinkmann AO, Deboth NJ, Trapman J. The mouse androgen receptor. Functional analysis of the protein and characterization of the gene. Biochem J. 1991; 278:269–278. [PubMed: 1883336]

Ferlin A, Bartoloni L, Rizzo G, Roverato A, Garolla A, Foresta C. Androgen receptor gene CAG and GGC repeat lengths in idiopathic male infertility. Mol Hum Reprod. 2004; 10:417–421. DOI: 10.1093/molehr/gah054 [PubMed: 15044606]

Garai C, Furuichi T, Kawamoto Y, Ryu H, Inoue-Murayama M. Androgen receptor and monoamine oxidase polymorphism in wild bonobos. Meta Gene. 2014; 2:831–343. DOI: 10.1016/j.mgene. 2014.10.005 [PubMed: 25606465]

Gelmann EP. Molecular biology of the androgen receptor. J Clin Oncol. 2002; 20:3001–3015. [PubMed: 12089231]

He WW, Fischer LM, Sun SQ, Bilhartz DL, Zhu XP, Young CYF, Kelley DB, Tindall DJ. Molecular cloning of androgen receptors from divergent species with a polymerase chain-reaction technique: complete cDNA sequence of the mouse androgen receptor and isolation of androgen receptor cDNA probes from dog, guinea pig and clawed frog. Biochem Biophys Res Commun. 1990; 171:697–704. DOI: 10.1016/0006-291x(90)91202-4 [PubMed: 2403358]

Hong KW, Hibino E, Takenaka O, Hayasaka I, Murayama Y, Ito S, Inoue-Murayama M. Comparison of androgen receptor CAG and GGN repeat length polymorphism in humans and apes. Primates. 2006; 47:248–254. DOI: 10.1007/s10329-005-0174-4 [PubMed: 16467955]

Hsing AW, Gao YT, Wu G, Wang X, Deng J, Chen YL, Sesterhenn IA, Mostofi FK, Benichou J, Chang C. Polymorphic CAG and GGN repeat lengths in the androgen receptor gene and prostate cancer risk: a population-based case-control study in China. Cancer Res. 2000; 60:5111–5116. [PubMed: 11016637]

Huhtaniemi IT, Pye SR, Limer KL, Thomson W, O'Neill TW, Platt H, Payne D, John SL, Jiang M, Boonen S, Borghs H, Vanderschueren D, Adams JE, Ward KA, Bartfai G, Casanueva F, Finn JD, Forti G, Giwercman A, Han TS, Kula K, Lean ME, Pendleton N, Punab M, Silman AJ, Wu FC. Increased estrogen rather than decreased androgen action is associated with longer androgen receptor CAG repeats. J Clin Endocrinol Metab. 2009; 94:277–284. DOI: 10.1210/jc.2008-0848 [PubMed: 18840639]

Inoue-Murayama M, Hibino E, Iwatsuki H, Inoue E, Hong KW, Nishida T, Hayasaka I, Ito S, Murayama Y. Interspecies and intraspecies variations in the serotonin transporter gene intron 3 VNTR in nonhuman primates. Primates. 2008; 49:139–142. DOI: 10.1007/s10329-007-0077-7 [PubMed: 18204817]

Jonsson EG, von Gertten C, Gustavsson JP, Yuan QP, Lindblad-Toh K, Forslund K, Rylander G, Mattila-Evenden M, Asberg M, Schalling M. Androgen receptor trinucleotide repeat polymorphism and personality traits. Psychiatr Genet. 2001; 11:19–23. DOI: 10.1097/00041444-200103000-00004 [PubMed: 11409695]

Konno A, Inoue-Murayama M, Hasegawa T. Androgen receptor gene polymorphisms are associated with aggression in Japanese Akita Inu. Biol Lett. 2011; 7:658–660. DOI: 10.1098/rsbl.2011.0087 [PubMed: 21450727]

Kuiper GGJM, Faber PW, Vanrooij HCJ, Vanderkorput JACM, Risstalpers C, Klaassen P, Trapman J, Brinkmann AO. Structural organization of the human androgen receptor gene. J Mol Endocrinol. 1989; 2:R1–R4. [PubMed: 2546571]

Li J, Al-Azzawi F. Mechanism of androgen receptor action. Maturitas. 2009; 63:142–148. DOI: 10.1016/j.maturitas.2009.03.008 [PubMed: 19372015]

Migeon BR, Brown TR, Axelman J, Migeon CJ. Studies of the locus for androgen receptor: localization on the human X chromosome and evidence for homology with the Tfm locus in the mouse. Proc Natl Acad Sci U S A. 1981; 78:6339–6343. DOI: 10.1073/pnas.78.10.6339 [PubMed: 6947233]

Mubiru JN, Cavazos N, Hemmat P, Garcia-Forey M, Shade RE, Rogers J. Androgen receptor CAG repeat polymorphism in males of six non-human primate species. J Med Primatol. 2012; 41:67–70. DOI: 10.1111/j.1600-0684.2011.00517.x [PubMed: 21978295]

Poux C, Madsen O, Marquard E, Vieites DR, de Jong WW, Vences M. Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores, and rodents as inferred from nuclear genes. Syst Biol. 2005; 54:719–730. DOI: 10.1080/10635150500234534 [PubMed: 16243759]

Quigley CA, De Bellis A, Marschke KB, El-Awady MK, Wilson EM, French FS. Androgen receptor defects: historical, clinical, and molecular perspectives. Endocr Rev. 1995; 16:271–321. DOI: 10.1210/edrv-16-3-271 [PubMed: 7671849]

Rajender S, Singh L, Thangaraj K. Phenotypic heterogeneity of mutations in androgen receptor gene. Asian J Androl. 2007; 9:147–179. DOI: 10.1111/j.1745-7262.2007.00250.x [PubMed: 17334586]

Rubinow DR, Schmidt PJ. Androgens, brain, and behavior. Am J Psychiatry. 1996; 153:974–984. [PubMed: 8678193]

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li SH, Margolis RL, Ross CA, Ferguson-Smith MA. Microsatellite evolution--evidence for directionality and variation in rate between species. Nat Genet. 1995; 10:337–343. DOI: 10.1038/ng0795-337 [PubMed: 7670473]

Santangelo AM, Ito M, Shiba Y, Clarke HF, Schut EHS, Cockcroft G, Ferguson-Smith AC, Roberts AC. Novel primate model of serotonin transporter genetic polymorphisms associated with gene expression, anxiety and sensitivity to antidepressants. Neuropsychopharmacology. 2016; 41:2366–2376. DOI: 10.1038/npp.2016.41 [PubMed: 26997299]

Schoof VAM, Jack KM. The association of intergroup encounters, dominance status, and fecal androgen and glucocorticoid profiles in wild male white-faced capuchins (*Cebus capucinus*). Am J Primatol. 2013; 75:107–115. DOI: 10.1002/ajp.22089 [PubMed: 23090872]

Snipes C, Forest M, Migeon C. Plasma androgen concentrations in several species of Old and New World monkeys. Endocrinology. 1969; 85:794–798. [PubMed: 4240879]

Spencer JA, Watson JM, Lubahn DB, Joseph DR, French FS, Wilson EM, Graves JAM. The androgen receptor gene is located on a highly conserved region of the X chromosomes of marsupial and monotreme as well as eutherian mammals. J Hered. 1991; 82:134–139. [PubMed: 2013687]

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28:2731–2739. DOI: 10.1093/molbev/msr121 [PubMed: 21546353]

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleicleic Acids Res. 1994; 22:4673–4680.
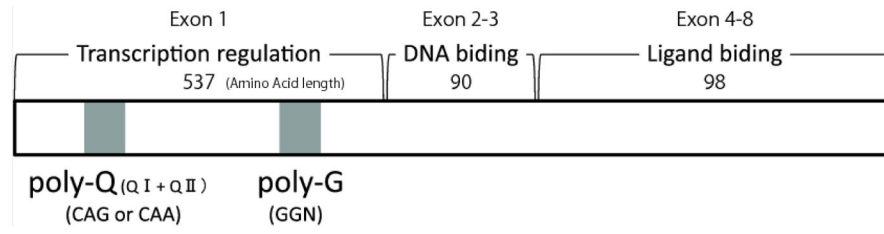
Fig. 1.
Schematic illustration of the structure of AR. An example amino acid length of human AR is shown (modified from Gelmann 2002).
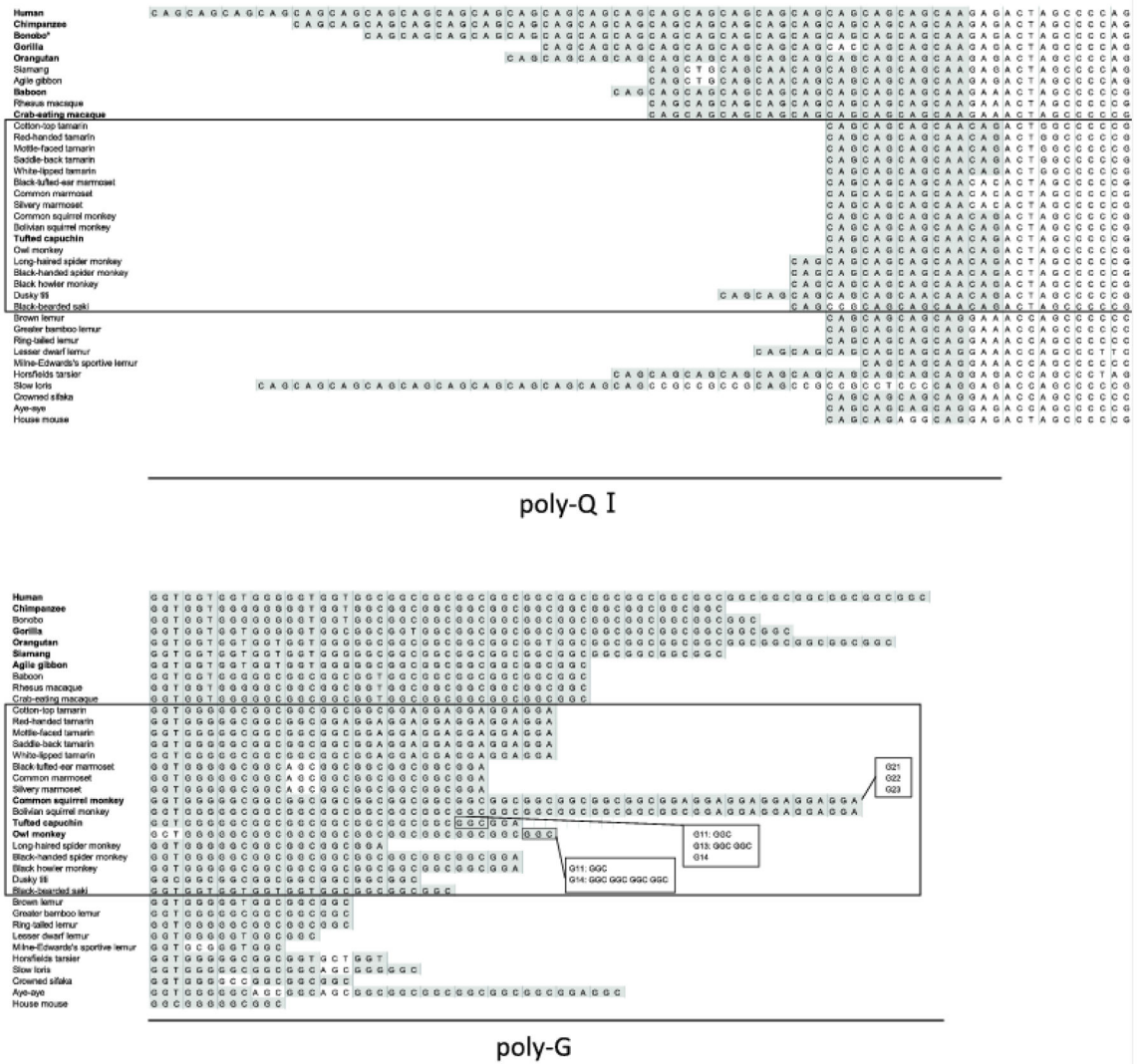
Fig. 2.
Alignment of nucleotide sequences of representative alleles in primates for poly-Q and poly-G regions. Glutamine and glycine codons are highlighted. The sequences of New World monkey species were obtained in this study. Names in bold indicate species with polymorphism. Polymorphic sites found in this study are enclosed with squares and sequence of other alleles are illustrated except common squirrel monkey G22 and G23, and tufted capuchin monkey G14 for which samples for sequencing analysis were not available. The sequences of other taxa were obtained from previous reports registered in DDBJ/EMBL/GenBank nucleotide sequence databases.
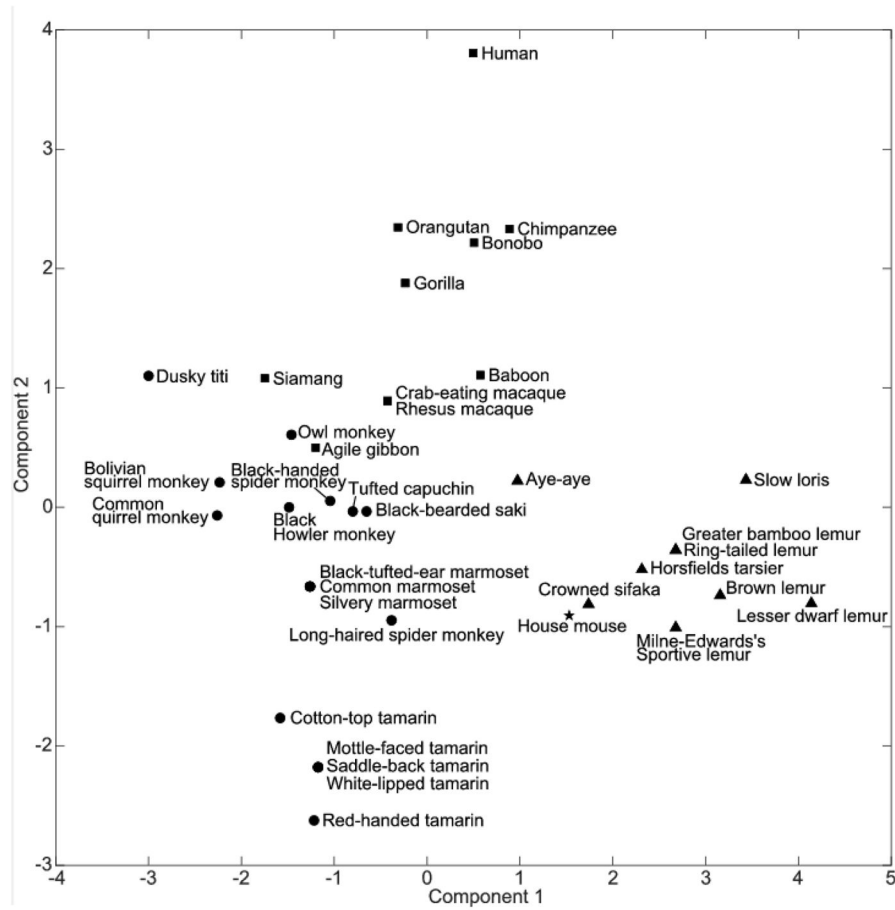
Fig. 3.
Distribution of representative alleles from each species in a two-dimensional MDS space based on the lengths of poly-QI, poly-QII and poly-G regions and the codon usage in these regions. Circle: New World monkeys; square: catarrhines; triangle: prosimians; star: non-primate (house mouse).
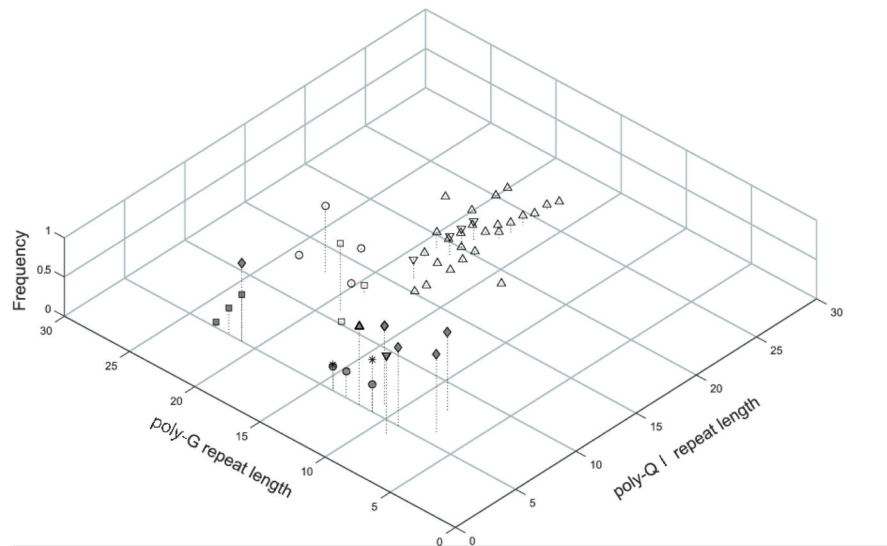
Fig. 4.

Relationship between poly-QI and poly-G length. Combinations of poly-QI and poly-G were obtained from hemizygous males and homozygous females on either region. Filled symbols indicate New World monkeys (filled circle: capuchin monkeys; filled square: squirrel monkeys; filled triangle: tamarin; filled inverse triangle: marmoset; asterisk: owl monkey; filled diamond: other species). Open symbols indicate apes (open circle: orangutan; open square: gorilla; open triangle; chimpanzee; open inverse triangle: bonobo). Note that frequency of each combination is represented as elevation of a dashed line (z-axis value) of a plot for species with polymorphism. Elevation was set to be 1 for monomorphic species.

Dataset analyzed in this study.

Table 1

taxon	species (scientific name)	# of individuals surveyed (reference)	accession numbers of sequences in Fig. 2
	Human (*Homo sapiens*)	304 (Hsing et al. 2000); 2878 (Huhtaniemi et al. 2009)	M20132 (Lubahn et al. 1988)
	Chimpanzee (*Pan troglodytes*)	57 (Hong et al. 2006); 48 (Mubiru et al. 2011)[*]	NM001009012 (Choong et al. 1998)
	Bonobo (*Pan paniscus*)	2 (Hong et al. 2006); 32 (Garai et al. 2014)	Q: AB970512; G: AB970515 (Garai et al. 2014)
	Gorilla (*Gorilla gorilla*)	18 (Hong et al. 2006); 3 (Djian et al. 1996)	Q: AB207220; G: AB207223 (Hong et al. 2006)
	Orangutan (*Pongo pygmaeus*)	20 (Hong et al. 2006); 4 (Murayama unpublished data)	Q: AB207232; G: AB207226 (Hong et al. 2006)
catarrhine	Siamang (*Symphalangus syndactylus*)	17 (Hong et al. 2006)	Q: AB207227; G: AB207229 (Hong et al. 2006)
	Agile gibbon (*Hylobates agilis*)	16 (Hong et al. 2006)	Q: AB207217; G: AB207216 (Hong et al. 2006)
	Baboon (*Papio hamadryas*)	6 (Choong et al. 1998); 230 (Mubiru et al. 2011)[*]	U94176 (Choong et al. 1998)
	Mangabey (*Cercocebus atys*)	56 (Mubiru et al. 2011)[*]	Sequences have not been reported in the reference.
	Rhesus macaque (*Macaca mulatta*)	54 (Mubiru et al. 2011)[*]	NM_001032911 (Chen et al. 2008)
	Crab-eating macaque (*Macaca fascicularis*)	6 (Choong et al. 1998); 23 (Mubiru et al. 2011)[*]	U94179 (Choong et al. 1998)
	Cotton-top tamarin (*Saguinus oedipus*)	23 (f=10, m=13)	Q: LC185460; G: LC185479
	Red-handed tamarin (*Saguinus midas*)	6 (f=4, m=2)	Q: LC185461; G: LC185480
	Mottle-faced tamarin (*Saguinus inustus*)	2 (f=0 , m=2)	Q: LC185464; G: LC185483
	Saddle-back tamarin (*Saguinus fuscicollis*)	2 (f=0, m=2)	Q: LC185463; G: LC185482
	White-lipped tamarin (*Saguinus labiatus*)	3 (f=1, m=2)	Q: LC185462; G: LC185481
	Black-tufted-ear marmoset (*Callithrix penicillata*)	6 (f=4 , m=2)	Q: LC185465; G: LC185484
New World monkey (platyrrhine)	Common marmoset (*Callithrix jacchus*)	1 (f=1 , m=0); 48 (Mubiru et al.2011)[*]	Q: LC185467; G: LC185486
	Silvery marmoset (*Callithrix argentata*)	1 (f=1, m=0)	Q: LC185466; G: LC185485
	Common squirrel monkey (*Saimiri sciureus*)	10 (f=5, m=5)	Q: LC185468; G: LC185487
	Bolivian squirrel monkey (*Saimiri boliviensis*)	6 (f=4, m=2)	Q: LC185469; G: LC185488
	Tufted capuchin (*Sapajus apella*)	47 (f=24, m=23)	Q: LC185470 (4 repeats), LC185471 (5 repeats); G: LC185489 (11 repeats), LC185490 (13 repeats)
	Owl monkey (*Aotus trivirgatus*)	5 (f=4, m=1)	Q: LC185472 (91Ser), LC185473 (91Gly); G: LC185491 (11 repeats), LC185492 (14 repeats)

taxon	species (scientific name)	# of individuals surveyed (reference)	accession numbers of sequences in Fig. 2
	Long-haired spider monkey (*Ateles belzebuth*)	3 (f=0, m=3)	Q: LC185474; G: LC185493
	Black-handed spider monkey (*Ateles geoffroyi*)	1 (f=0, m=1)	Q: LC185475; G: LC185494
	Black howler monkey (*Alouatta caraya*)	2 (f=2, m=0)	Q: LC185476; G: LC185495
	Dusky titi (*Plecturocebus moloch*)	1 (f=1, m=0)	Q: LC185477; G: LC185496
	Black-bearded saki (*Chiropotes satanas*)	1 (f=1, m=0)	Q: LC185478; G: LC186017
	Brown lemur (*Eulemur fulvus*)	1 (Choong et al. 1998); (Poux et al. 2005)1	U94178 (Choong et al. 1998)
	Greater bamboo lemur (*Prolemur simus*)	1 (Poux et al. 2005)	AJ893538 (Poux et al. 2005)
	Ring-tailed lemur (*Lemur catta*)	1 (Poux et al. 2005)	AJ893536 (Poux et al. 2005)
	Lesser dwarf lemur (*Cheirogaleus medius*)	1 (Poux et al. 2005)	AJ893540 (Poux et al. 2005)
prosimian	Milne-Edwards's Sportive lemur (*Lepilemur edwardsi*)	1 (Poux et al. 2005)	AJ893539 (Poux et al. 2005)
	Horsfields tarsier (*Cephalopachus bancanus*)	1 (Poux et al. 2005)	AJ893544 (Poux et al. 2005)
	Slow loris (*Mycticebus coucang*)	1 (Poux et al. 2005)	AJ893543 (Poux et al. 2005)
	Crowned sifaka (*Propithecus coronatus*)	1 (Poux et al. 2005)	AJ893542 (Poux et al. 2005)
	Aye-aye (*Daubentonia madagascariensis*)	1 (Poux et al. 2005)	AJ893541 (Poux et al. 2005)
non-primate	House mouse (*Mus musculus*)	1 (Faber et al. 1991)	X59592 (Faber et al. 1991)

All data of New World monkeys (platyrrhine) were obtained in this study except for 48 samples of common marmosets. Data for other taxa were obtained from references indicated in parenthesis. Number of female and male individuals surveyed in this study is indicated in parenthesis. Bold letters indicate predominant codon in each species. Accession numbers at DDBJ/EMBL/GenBank nucleotide sequence database for sequences in Fig. 2 are provided.

*
Only polymorphism of poly-Q region was surveyed in Mubiru et al. 2012.

Table 2

Repeat length of AR gene poly-Q and poly-G regions in primates.

taxon	species (scientific name)	Q repeats (QI)*	Q repeats (QII)	G repeats
		range (frequent allele)	range (frequent allele)	range (frequent allele)
catarrhine	Human (*Homo sapience*)	10–37 (23)	6	14–27 (23)
	Chimpanzee (*Pan troglodytes*)	15–27 (19)	5	14–22 (17)
	Bonobo (*Pan paniscus*)	12–21(17)	5	18–19 (18)
	Gorilla (*Gorilla gorilla*)	7–18 (11)	5	18–19 (19)
	Orangutan (*Pongo pygmaeus*)	13–16 (13)	5	20–24 (22)
	Siamang (*Symphalangus syndactylus*)	8	5	16–21 (17)
	Agile gibbon (*Hylobates agilis*)	8	5	11–13 (13)
	Baboon (*Papio hamadryas*)	5–14 (10)	6	13
	Mangabey (*Cercocebus atys*)	9–11 (9)	8*	-
	Rhesus macaque (*Macaca mulatta*)	9	7	13
	Crab-eating macaque (*Macaca fascicularis*)	8–10 (9)	7	13
New World monkey (platyrrhine)	Cotton-top tamarin (*Saguinus oedipus*)	5	4	12
	Red-handed tamarin (*Saguinus midas*)	5	5	12
	Mottle-faced tamarin (*Saguinus inustus*)	5	5	12
	Saddle-back tamarin (*Saguinus fuscicollis*)	5	5	12
	White-lipped tamarin (*Saguinus labiatus*)	5	5	12
	Black-tufted-ear marmoset (*Callithrix penicillata*)	4	4	9
	Common marmoset (*Callithrix jacchus*)	4	4	9
	Silvery marmoset (*Callithrix argentata*)	4	4	9
	Common squirrel monkey (*Saimiri sciureus*)	5	4	21–23 (21)
	Bolivian squirrel monkey (*Saimiri boliviensis*)	5	4	21
	Tufted capuchin (*Sapajus apella*)	5	4–5 (5)	11–14 (11)
	Owl monkey (*Aotus trivirgatus*)	5	4	11–14 (11)
	Long-haired spider monkey (*Ateles belzebuth*)	6	4	7

taxon	species (scientific name)	# of repeat		
		Q repeats (QI)[*] range (frequent allele)	Q repeats (QII) range (frequent allele)	G repeats range (frequent allele)
	Black-handed spider monkey (*Ateles geoffroyi*)	6	4	11
	Black howler monkey (*Alouatta caraya*)	6	3	11
	Dusky titi (*Plecturocebus moloch*)	8	4	8
	Black-bearded saki (*Chiropotes satanas*)	5	4	9
	Brown lemur (*Eulemur fulvus*)	4	7	6
	Greater bamboo lemur (*Prolemur simus*)	4	7	6
	Ring-tailed lemur (*Lemur catta*)	4	7	6
prosimian	Lesser dwarf lemur (*Cheirogaleus medius*)	6	9	5
	Milne-Edwards's Sportive lemur (*Lepilemur edwardsi*)	3	5	3
	Horsfields tarsier (*Cephalopachus bancanus*)	10	6	6
	Slow loris (*Mycticebus coucang*)	13	7	7
	Crowned sifaka (*Propithecus coronatus*)	4	8	5
	Aye-aye (*Daubentonia madagascariensis*)	4	5	12
non-primate	House mouse (*Mus musculus*)	3	2	4

For species with length polymorphisms, the range of length and the most frequent allele (in parenthesis) are indicated.

[*]
Number was obtained from the sequence in accession No. XM_012060885.

Table 3

Allele frequencies of the poly-QII and poly-G repeats in species with polymorphism detected in this study.

species		common squirrel monkey			tafted capuchin			owl monkey		
sample size		f=5, m=5, n=15			f=24, m=23, n=71			f=4, m=1, n=9		
	allele	frequency	heterozygosity in females	allele	frequency	heterozygosity in females	allele	frequency	heterozygosity in females
poly-QII repeat				QII4	0.48	He: 0.50			
				QII5	0.52	He: 0.63			
poly-G repeat	G21	0.60		G11	0.35		G11	0.67	
	G22	0.33	He: 0.52 Ho: 1.00	G13	0.34	He: 0.63 Ho: 0.71	G14	0.33	He: 0.44 Ho: 1.00
	G23	0.07		G14	0.31				

The expected heterozygosity (He), and the observed heterozygosity (Ho) in females are also provided. :n*' indicates number of chromosomes.

Table 4

Codon usage of AR gene poly-Q and poly-G regions in primates.

taxon	species (scientific name)	Q region		G region			
		CAA	CAG	GGA	GGT	GGG	GGC
catarrhine	Human (*Homo sapience*)	0.03	1	0	0.22	0.04	0.7
	Chimpanzee (*Pan troglodytes*)	0.04	1	0	0.24	0.12	0.7
	Bonobo (*Pan paniscus*)	0.05	1	0	0.22	0.11	0.7
	Gorilla (*Gorilla gorilla*)	0.06	0.9	0	0.21	0.05	0.7
	Orangutan (*Pongo pygmaeus*)	0.06	0.9	0	0.23	0.05	0.7
	Siamang (*Symphalangus syndactylus*)	0.15	0.9	0	0.24	0.06	0.7
	Agile gibbon (*Hylobates agilis*)	0.15	0.9	0	0.31	0.08	0.6
	Baboon (*Papio hamadryas*)	0.06	0.9	0	0.23	0.08	0.7
	Rhesus macaque (*Macaca mulatta*)	0.13	0.9	0	0.23	0.08	0.7
	Crab-eating macaque (*Macaca fascicularis*)	0.13	0.9	0	0.23	0.08	0.7
New World monkey (platyrrhine)	Cotton-top tamarin (*Saguinus oedipus*)	0.11	0.9	0.4	0.08	0.08	0.4
	Red-handed tamarin (*Saguinus midas*)	0.1	0.9	0.6	0.08	0.08	0.25
	Mottle-faced tamarin (*Saguinus inustus*)	0.1	0.9	0.5	0.08	0.08	0.33
	Saddle-back tamarin (*Saguinus fuscicollis*)	0.1	0.9	0.5	0.08	0.08	0.33
	White-lipped tamarin (*Saguinus labiatus*)	0.1	0.9	0.5	0.08	0.08	0.33
	Black-tufted-ear marmoset (*Callithrix penicillata*)	0.13	0.9	0.11	0.11	0.11	0.7
	Common marmoset (*Callithrix jacchus*)	0.13	0.9	0.11	0.11	0.11	0.7
	Silvery marmoset (*Callithrix argentata*)	0.13	0.9	0.11	0.11	0.11	0.7
	Common squirrel monkey (*Saimiri sciureus*)	0.11	0.9	0.29	0.05	0.05	0.6
	Bolivian squirrel monkey (*Saimiri boliviensis*)	0.11	0.9	0.24	0.05	0.05	0.7
	Tufted capuchin (*Sapajus apella*)	0.1	0.9	0.09	0.09	0.09	0.7
	Owl monkey (*Aotus trivirgatus*)	0.11	0.9	0	0	0.09	0.9
	Long-haired spider monkey (*Ateles belzebuth*)	0.1	0.9	0.14	0.14	0.14	0.6
	Black-handed spider monkey (*Ateles geoffroyi*)	0.1	0.9	0.09	0.09	0.09	0.7
	Black howler monkey (*Alouatta caraya*)	0.11	0.9	0.09	0.09	0.09	0.7

taxon	species (scientific name)	codon usage					
		Q region		G region			
		CAA	CAG	GGA	GGT	GGG	GGC
	Dusky titi (*Plecturocebus moloch*)	0.17	0.8	0	0	0	1
	Black-bearded saki (*Chiropotes satanas*)	0.11	0.9	0	0.44	0	0.6
	Brown lemur (*Eulemur fulvus*)	0	1	0	0.33	0.17	0.5
	Greater bamboo lemur (*Prolemur simus*)	0	1	0	0.17	0.17	0.7
	Ring-tailed lemur (*Lemur catta*)	0	1	0	0.17	0.17	0.7
prosimian	Lesser dwarf lemur (*Cheirogaleus medius*)	0	1	0	0.4	0.2	0.4
	Milne-Edwards's Sportive lemur (*Lepilemur edwardsi*)	0	1	0	0.7	0	0.33
	Horsfields tarsier (*Cephalopachus bancanus*)	0.06	0.9	0	0.5	0.17	0.33
	Slow loris (*Nycticebus coucang*)	0	1	0	0.14	0.29	0.6
	Crowned sifaka (*Propithecus coronatus*)	0.08	0.9	0	0.2	0.2	0.6
	Aye-aye (*Daubentonia madagascariensis*)	0	1	0.08	0.08	0.08	0.8
non-primate	House mouse (*Mus musculus*)	0	1	0	0	0.25	0.8