# An Explicit Signature of Balancing Selection for Color-Vision Variation in New World Monkeys

Tomohide Hiwatashi,[1] Yugo Okabe,[1] Toko Tsutsui,[1] Chihiro Hiramatsu,[†,1] Amanda D. Melin,[2] Hiroki Oota,[1] Colleen M. Schaffner,[3] Filippo Aureli,[4] Linda M. Fedigan,[2] Hideki Innan,[5] and Shoji Kawamura[*,1]

[1]Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan

[2]Department of Anthropology, University of Calgary, Calgary, AB, Canada

[3]Psychology Department, University of Chester, Chester, United Kingdom

[4]Research Centre in Evolutionary Anthropology and Palaeoecology, School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, United Kingdom

[5]Department of Evolutionary Studies and Biosystems, The Graduate University for Advanced Studies (Sokendai), Hayama, Japan

†Present address: Department of Information Physiology, Division of Sensory and Cognitive Information, National Institute for Physiological Science, Okazaki 444-8585, Japan

***Corresponding author:** E-mail: Kawamura@k.u-tokyo.ac.jp.

**Associate editor:** Naoko Takezaki

Research article

## Abstract

Color vision is an important characteristic of primates and, intriguingly, Neotropical monkeys are highly polymorphic for this trait. Recent field studies have challenged the conventional view that trichromatic color vision is more adaptive than dichromatic color vision. No study has investigated the pattern of genetic variation in the long to middle wavelength–sensitive (L-M or red–green) opsin gene as compared with that of other genomic regions (neutral references) in wild populations of New World monkeys to look for the signature of natural selection. Here, we report such a study conducted on spider monkeys and capuchin monkeys inhabiting Santa Rosa National Park, Costa Rica. The nucleotide sequence of the L-M opsin gene was more polymorphic than the sequences of the neutral references, although the opsin-gene sequences were not more divergent between the two species than were the sequences of the neutral references. In a coalescence simulation that took into account the observed nucleotide diversity of the neutral references, the Tajima's *D* value of the L-M opsin gene deviated significantly in a positive direction from the expected range. These results are the first to statistically demonstrate balancing selection acting on the polymorphic L-M opsin gene of New World monkeys. Taking the results of behavioral and genetic studies together, the balancing selection we detected may indicate that coexistence of different color-vision types in the same population, also characteristic of humans, is adaptive.

**Key words:** opsin, color-vision polymorphism, balancing selection, New World monkeys, *Ateles geoffroyi*, *Cebus capucinus*.

## Introduction

Among placental mammals, only primates have evolved trichromatic color vision by having two subtypes of the X-chromosomal long to middle wavelength–sensitive (L-M) opsin gene in addition to an autosomal short wavelength sensitive (S) opsin gene. The two subtypes occur via allelic differentiation (e.g., most New World monkeys) or gene duplication (e.g., Old World monkeys, apes and humans) of the L-M opsin gene, enabling enhanced resolution of red–green color contrast (Jacobs 1996). The selective advantage of trichromacy has been supported by many theoretical studies of object visibility based on the colorimetric properties of natural scenes measured in forests (Regan et al. 1998, 2001; Sumner and Mollon 2000, 2003; Parraga et al. 2002). Other colorimetric studies incorporating nutritional measurements of primate diets also support a trichromat advantage in foraging for young leaves or fruits (Lucas et al. 1998, 2003; Dominy and Lucas 2001, 2004; Riba-Hernandez et al. 2005). Standardized behavioral experiments have demonstrated the superior ability of trichromatic as compared with dichromatic primates in detecting reddish objects against a greenish background (Caine and Mundy 2000; Caine 2002; Smith, Buchanan-Smith, Surridge, Osorio, and Mundy 2003). Interspecies comparisons of the L-M opsin-gene sequences have also found signatures of positive natural selection for generating the trichromatic color vision and have identified amino acid substitutions relevant for spectral differentiation between the L and M opsin genes (Yokoyama R and Yokoyama S 1990; Neitz et al. 1991; Shyue et al. 1995, 1998; Boissinot et al. 1998).

New World monkeys are an excellent model to test the suggested advantage of trichromacy because of the allelic polymorphism of the L-M opsin gene that results in coexistence of dichromatic and trichromatic individuals in the same population (Mollon et al. 1984). The color-vision polymorphism is transspecific and documented in all three families of the New World monkeys (Atelidae, Pitheciidae, and Cebidae) (Jacobs 2007). The long duration of the polymorphism in these Neotropical primate families is

consistent with balancing selection, a form of positive natural selection, operating to maintain variation via heterozygote advantage of trichromatic females (Boissinot et al. 1998; Surridge and Mundy 2002; Surridge et al. 2003).

However, despite these findings, many behavioral observations of wild primates have produced equivocal or contrary results to the pattern expected from the trichromat advantage hypothesis. A study of wild mixed-species troops of saddleback (*Saguinus fuscicollis*) and mustached (*Saguinus mystax*) tamarins showed that neither the color-vision types (dichromatic or trichromatic) nor the sex of individuals had a consistent effect on the leadership of the troops to feeding trees (Smith, Buchanan-Smith, Surridge, and Mundy 2003). Another study of tamarins (*Saguinus imperator imperator* and *Saguinus fuscicollis weddelli*) found no significant difference between females (thought to consist of trichromats and dichromats) and males (all dichromats) in their ability to locate or discriminate feeding sites (Dominy et al. 2003). In a population of capuchin monkeys (*Cebus capucinus*), no significant difference between trichromats and dichromats was found in feeding and energy intake rates (Vogel et al. 2007). In another population of the same capuchin monkey species, there was no difference in foraging time spent on different food types (Melin et al. 2008). Some modeling studies have found that many fruits eaten by spider monkeys (*Ateles geoffroyi*) or squirrel monkeys (*Saimiri sciureus*) are similarly discernible or similarly indiscernible from background foliage for both trichromats and dichromats (Riba-Hernandez et al. 2004; Stoner et al. 2005; De Araujo et al. 2006). A field study of free-ranging spider monkeys (*A. geoffroyi*) measuring their foraging efficiency on fruits and colorimetric properties of fruits and background leaves revealed that dichromats are not inferior to trichromats in frequency, accuracy, and unit-time intake efficiency of detecting fruits and that this is because the luminance contrast of fruits to background leaves is the main determinant of the fruit detection in both dichromats and trichromats (Hiramatsu et al. 2008). A study of the same social group of spider monkeys also showed that irrespective of color-vision phenotypes the monkeys sniff visually cryptic fruits more often than visually conspicuous fruits and that olfactory inspection was an important determinant for ingestion or rejection of fruits (Hiramatsu et al. 2009). Behavioral experiments using capuchin monkeys (*Cebus apella*) and marmosets (*Callithrix geoffroyi*) to detect color-camouflaged objects (Caine et al. 2003; Saito, Mikami, et al. 2005) have even suggested a trichromat disadvantage and a field study of capuchin monkeys (*C. capucinus*) has demonstrated a dichromat advantage in foraging for surface-dwelling insects (Melin et al. 2007). These behavioral observations suggest that the superior ability of trichromats to see red–green color contrast may not translate into a selective advantage because the use of a variety of sensory modalities may compensate for the inferiority of any one sense (Hiramatsu et al. 2009).

The effective population size $N_e$, a major determinant of the duration of allelic turnover, remains unknown for New World monkeys. Assuming that the last common ancestor of all New World monkeys originated 26 Ma (Schneider 2000), it follows that the opsin polymorphism has persisted over this period. In theory, the expected survival time for a neutral X linked allele is $3N_e$ generations in a stationary population. If $N_e$ of New World monkeys had been large enough (e.g., in the order of $10^6$) in the long isolated South American continent without formidable eutherian predators, then the polymorphism could have persisted for this length of time. Another problem is that it is difficult to estimate $N_e$ in a natural population. The estimated value of $N_e$ depends on the accuracy of estimates of the mutation rate and generation time and is confounded by demographic effects such as the historical dynamism of population size, migration pattern, and population structure. Although demographic effects influence genetic variation of all genes in the genome alike, the pattern and the intensity of natural selection can vary among genomic regions depending on direct or indirect effects of mutations in that region to fitness. We thus applied a method that compared the pattern of intraspecific genetic variation between a focal region (i.e., the L-M opsin gene) and other reference regions in the same genome using the same-population samples to cancel out the effects of demographic factors that both regions share (Verrelli and Tishkoff 2004; Perry et al. 2007; Verrelli et al. 2008). In this study, we employed a computer-simulation approach to determine the statistical significance of difference in polymorphism level between the L-M opsin gene and reference regions that are presumably evolving neutrally and sought evidence for positive natural selection on the L-M opsin gene in wild populations of New World monkeys.

## Materials and Methods

### Genotyping of the L-M Opsin Gene

The L-M opsin gene consists of six exons, encoding a protein of 364 amino acids long and spanning approximately 15 kb in the genome in primates (Nathans et al. 1986; Kawamura et al. 2001; Nagao et al. 2005). The mutagenesis studies have shown that the peak absorption spectra ($\lambda_{max}$) of the vertebrate M/LWS class of opsins, where the primate L-M opsins belong, can be predicted from the amino acid composition at the three sites, 180, 277, and 285, together with two additional sites, 197 and 308 ("five-site rule") (Sun et al. 1997; Yokoyama and Radlwimmer 1998, 2001; Yokoyama et al. 2008). Among primate L-M opsins, residues 197 and 308 are not varied and are irrelevant to spectral differences among them. Therefore, the five-site rule can be reduced to the "three-site rule" in primates, where amino acid changes from Ser to Ala at site 180 (denoted Ser180Ala), Tyr277Phe, and Thr285Ala shift the $\lambda_{max}$ values by −7, −8, and −15 nm, respectively, and the reverse amino acid changes cause opposite spectral shifts by the same extent in a nearly additive manner (Yokoyama and Radlwimmer 2001; Hiramatsu et al. 2004). The residue 180 is encoded in exon 3 and 277 and 285 are in exon 5. Thus, the spectral genotype of the L-M opsin gene can be determined through polymerase chain reaction (PCR)

**Table 1.** Genotype and Allele Frequencies of the L-M Opsin Gene in Spider and Capuchin Monkeys.

| Species | Phenotype | Genotype | No. of Females | No. of Males | Allele | Frequency |
|---|---|---|---|---|---|---|
| Spider monkeys | Trichromat | P553/P538 | 9 | n/a | P553 | 0.596 |
| | | | | | P538 | 0.404 |
| | Dichromat | P553/P553 (female), P553 (male) | 7 | 8 | | |
| | | P538/P538 (female), P538 (male) | 4 | 4 | | |
| | | Total | 20 | 12 | | |
| Capuchin monkeys | Trichromat | P561/P543 | 3 | n/a | P561 | 0.516 |
| | | P543/P532 | 2 | n/a | P543 | 0.355 |
| | | P561/P532 | 1 | n/a | P532 | 0.129 |
| | Dichromat | P561/P561 (female), P561 (male) | 1 | 10 | | |
| | | P543/P543 (female), P543 (male) | 1 | 4 | | |
| | | P532/P532 (female), P532 (male) | 0 | 1 | | |
| | | Total | 8 | 15 | | |

and DNA sequencing of fecal DNA for the two exons as previously described (Hiramatsu et al. 2004, 2005).

On the basis of the three-site composition, we previously identified three L-M opsin alleles in capuchin monkeys, one having the three-site composition Ser, Tyr, and Thr at sites 180, 277, and 285, respectively (designated Ser/Tyr/Thr) and the other two having Ala/Phe/Thr and Ala/Phe/Ala (Hiramatsu et al. 2005). The $\lambda_{max}$ values of the three alleles were directly measured by the method of photopigment reconstitution in vitro and were determined to be 561, 543, and 532 nm (designated P561, P543, and P532), respectively (Hiramatsu et al. 2005). Likewise, we previously identified two L-M opsin alleles in spider monkeys, one having the three-site composition Ser/Tyr/Thr and the other having Ser/Phe/Thr (Hiramatsu et al. 2005). The $\lambda_{max}$ values of the two alleles were directly measured in vitro and were determined to be 553 and 538 nm, respectively (designated P553 and P538) (Hiramatsu et al. 2008). With a single S opsin gene, female monkeys having two different spectral alleles of the L-M opsin gene are considered to be trichromats and females having two identical L-M opsin alleles and males, due to the hemizygosity of the X chromosome, are considered to be dichromats. Consistency between the genotype and phenotype has been well demonstrated by behavioral experiments (Caine and Mundy 2000; Saito, Kawamura, et al. 2005).

## Study Animals and Their Spectral Genotypes of the L-M Opsin Gene

We previously determined the spectral genotype of the L-M opsin gene using fecal DNA samples for 20 black-handed spider monkeys (*A. geoffroyi*) (17 females and 3 males) from one free-ranging social group (Santa Rosa 1 "SR1") and 19 white-faced capuchin monkeys (*C. capucinus*) (6 females and 13 males) from another (Los Valles "LV"), both inhabiting Santa Rosa National Park, Costa Rica (Hiramatsu et al. 2005). Spider and capuchin monkeys differ in diet (frugivores and omnivores, respectively), social structure (male philopatry with high fission–fusion dynamics and female philopatry with rather cohesive groups, respectively) (Fragaszy et al. 2004; Aureli and Schaffner 2008) and phylogeny (Atelidae and Cebidae, respectively) (Schneider 2000). For the present study, we collected more fecal samples from these groups and genotyped a total of 32 spider monkeys (20 females and 12 males) and 23 capuchin monkeys (8 females and 15 males) by following Hiramatsu et al. (2005) (table 1).

## Nucleotide Sequencing of the L-M Opsin-Gene Regions

All of the six protein-coding exons in spider monkey and four exons (1, 3, 5, and 6) in capuchin monkeys were PCR amplified from fecal DNA using primers listed in table 2 and were sequenced with their adjacent noncoding regions. These regions were named after the exons contained, with a capital for the first letter (e.g., "Exon 1": see table 3). An additional part of intron 4 was also included in the analysis for spider monkeys (table 2). The sequence lengths of these PCR products used in the analyses are summarized in table 3 with information on the length of exons and their adjacent noncoding regions.

The PCR was carried out in 50 μl containing 1.5 units of high fidelity Pyrobest polymerase (Takara, Tokyo, Japan) with 1× Pyrobest Buffer, 0.2 mM each of deoxyribonucleoside triphosphates (dNTPs), 1 μM each of the forward and reverse primers, and 5 μl of the DNA extract from the feces. Pure water was used as the template for negative control in every reaction. We carried out PCR at 94 °C for 2 min followed by 40 cycles at 98 °C for 10 s, 67 °C for 30 s, and 72 °C for 2 min. The amplified DNA fragments were purified by using UltraClean 15 (MO BIO Laboratories, Carlsbad, CA) from agarose gels. The purified DNA samples were directly sequenced using Applied Biosystems model 3130 automatic sequencer with Big Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems Japan, Tokyo) and with the PCR primers and, when necessary, other primers designed inside the amplified regions. For each spectral allele type, one male sample was subjected to complete sequencing of both strands (designated standard samples). For other samples, when a different nucleotide from the standard samples was found in two or more individuals at the same site repeatedly, we sequenced at least one of the samples in both strands to confirm the variation. When we found a different nucleotide in only one sample, that is, as a singleton among the samples, we repeated the PCR for the sample and sequenced the region for both strands to rule out a PCR error and confirm the variation.

**Table 2.** PCR Primer Pairs Used to Aamplify the L-M Opsin-Gene Regions and Neutral References in Spider and Capuchin Monkeys.

| Region | | Forward (5′ to 3′) | Reverse (5′ to 3′) |
|---|---|---|---|
| Spider monkey L-M opsin | Exon 1 | CTTGGCCCTGAAAAATCTCCCTG | CTGCTCCAAATGGAAAGACAGCC |
| | Exon 2 | TCTATGGAAGGGCAGAGGACTCT | TGTCTCCCCCAGGAAAAGCAGAT |
| | Exon 3 | CGTGAGTTGACCAGTGACAAGG | ACTGTCCAGTGCCCGCCTTTG |
| | Exon 4 | TCTCACTATGTGGTCCAGGCTAG | TGGTACTGCCCCAAAGGAATCAC |
| | Intron 4 | CAGCTCTGTGGGATAGGAAAGGA | ATGGCAGTGCCAGAGCAATGCAT |
| | Exon 5 | GCCATCTCCATCTTGTTAGCTCC | ATGTACCTAGGGCTCACCAACTC |
| | Exon 6 | TTGGGGAACACACTTCAACCCAG | CTCTCACTGACCATCGTCCTCTT |
| Capuchin monkey L-M opsin | Exon 1 | Same as spider monkeys | Same as spider monkeys |
| | Exon 3 | CTGATTCCTTGCTCTTGGCT | ACATTCCCTCTCCAAACACC |
| | Exon 5 | TCCCTCTCTCATCCCCACTCA | TGGATGTACCTAGGGCTCACC |
| | Exon 6 | Same as spider monkey | Same as spider monkey |
| η-Globin pseudogene | | AACTGCACATGGTGGAGGTAGTG | ACTCACCAGGAAGTTCTCAGGGT |
| vWF intron 11 | | AGCCCCTGGAAGACATGGAATGT | CAAGAGGGAACACAGGGTGACTT |
| TF intron 5 | | CCATGTTCCCTTTCATCTGACCC | CAGGGCTCCTGCAGGGCAAA |
| B2m introns 2 to 3 | | GCACGAATACATACGCACTCCTGAC | CACACCTGGGGCCACATACCT |
| S opsin intron 4 | | CCCTGCCAACTTTTAGCTTGCAC | TTCCCGCACACCATCTCCATGAT |

We did not attempt to separate two allelic sequences from females by DNA cloning and hence left the haplotype phases undetermined because phase information is not necessary for calculating $\pi$, $\theta_w$, Tajima's $D$ and interspecie divergence values (see the following sections). Thus, in the alignment of the nucleotide sequences (Supplementary Material online) used for calculating these values, two allele sequences of heterozygous females were only tentatively sorted into the allele sequences that were found in males and homozygous females. We excluded from the subsequent analyses the insertion–deletion sites and repeat elements, such as short interspersed nuclear elements and long interspersed nuclear elements, which we searched for using RepeatMasker (The Institute for Systems Biology, Seattle, WA: http://www.repeatmasker.org/), to avoid potential complexity.

**Table 3.** Lengths of the L-M Opsin-Gene Regions Used for the Analysis.

| | | Analyzed Length (bp) | |
|---|---|---|---|
| Region | Subregion | Spider Monkeys | Capuchin Monkeys |
| Exon 1 | 5′ flanking | 406 | 378 |
| | exon 1 | 112 | 112 |
| | intron 1 | 422 | 536 |
| | Total | 940 | 1,026 |
| Exon 2 | intron 1 | 371 | — |
| | exon 2 | 297 | — |
| | intron 2 | 347 | — |
| | Total | 1,015 | — |
| Exon 3 | intron 2 | 386 | 494 |
| | exon 3 | 169 | 169 |
| | intron 3 | 466 | 409 |
| | Total | 1,021 | 1,072 |
| Exon 4 | intron 3 | 147 | — |
| | exon 4 | 166 | — |
| | intron 4 | 152 | — |
| | Total | 465 | — |
| Intron 4 | | 827 | — |
| Exon 5 | intron 4 | 313 | 510 |
| | exon 5 | 240 | 240 |
| | intron 5 | 321 | 274 |
| | Total | 874 | 1,024 |
| Exon 6 | intron 5 | 157 | 139 |
| | exon 6 | 111 | 111 |
| | 3′ flanking | 631 | 586 |
| | Total | 899 | 836 |

## Nucleotide Sequencing of the Neutral References

The η-globin gene is suitable as a neutral reference because it is a single-copy pseudogene well characterized for primates (Koop et al. 1986; Bailey et al. 1991), and the ortholog of a spider monkey (*A. geoffroyi*) has been reported (GenBank accession number J03050) (Fitch et al. 1988). A pair of PCR primers (table 2) was designed according to the reported sequence of the spider monkey to amplify a region that is free of repeat sequences. Because we could not find other suitable pseudogenes of New World monkeys in public databases that are single copy and free of repeat sequences, we retrieved intron sequences of functional single-copy genes reported for New World monkeys: the von Willebrand factor (vWF) gene intron 11 (*A. geoffroyi*, GenBank DQ078118); the transferrin (TF) gene intron 5 (*A. geoffroyi*, GenBank DQ083094); the β-2-microglobulin precursor (B2m) gene introns 2–3 including 28 bp of exon 3 (*Ateles paniscus*, GenBank AF032087) (Canavez et al. 1998); the S opsin gene intron 4 (*Cebus olivaceus*, GenBank AF039424) (Shimmin et al. 1998). According to these sequences, PCR primer sets (table 2) were designed so as not to contain repeat sequences, which we searched for using RepeatMasker. The sequence lengths of these PCR products used in the analyses are indicated in table 4.

The PCR was carried out with the same composition of the reaction mixture as in the L-M opsin gene. Pure water was used as the template for negative control in every reaction. We carried out PCR at 94 °C for 2 min followed by 40 cycles at 98 °C for 10 s, 67 °C for 30 s, and 72 °C for 85 s. Purification and sequencing were carried out as in the L-M opsin gene. In both spider and capuchin monkeys, one sample each ("standard sample") was subjected to complete sequencing of both strands using the PCR and internal primers. For other samples, when a different nucleotide from the standard samples was found in two or more individuals at the same site repeatedly, we sequenced at least

**Table 4.** Summary Statistics of Nucleotide Variation in the Neutral References of Spider and Capuchin Monkeys.

| Species | Region | No. of Chromosomes Examined ($n$) | No. of Segregating Sites ($S$) | $\pi$ ($\times 10^{-3}$) | $P^a$ | $\theta_w$ ($\times 10^{-3}$) | $P^b$ | Tajima's $D$ | $P^c$ |
|---|---|---|---|---|---|---|---|---|---|
| Spider monkeys | η-Globin pseudogene (529 bp) | 64 | 4 | 1.36 | 0.7 | 1.60 | 0.7 | −0.32 | 0.7 |
| | vWF intron 11 (582 bp) | 64 | 7 | 3.08 | 0.1 | 2.54 | 0.2 | 0.53 | 0.2 |
| | TF intron 5 (551 bp) | 64 | 3 | 1.29 | 0.8 | 1.15 | 0.9 | 0.24 | 0.3 |
| | B2m introns 2 and 3 (537 bp) | 64 | 0 | 0.00 | 1 | 0.00 | 1 | 0.00 | 0.5 |
| | S opsin intron 4 (522 bp) | 64 | 5 | 3.03 | 0.1 | 2.03 | 0.4 | 1.14* | 0.03 |
| Capuchin monkeys | η-Globin pseudogene (500 bp) | 46 | 2 | 1.83 | 0.1 | 0.91 | 0.5 | 1.83** | 0.01 |
| | vWF intron 11 (500 bp) | 46 | 1 | 0.18 | 1 | 0.46 | 0.9 | −0.85 | 0.8 |
| | TF intron 5 (500 bp) | 46 | 0 | 0.00 | 1 | 0.00 | 1 | 0.00 | 0.5 |
| | B2m introns 2 and 3 (500 bp) | 42 | 2 | 1.06 | 0.4 | 0.93 | 0.5 | 0.25 | 0.4 |
| | S opsin intron 4 (500 bp) | 46 | 1 | 0.93 | 0.5 | 0.46 | 0.9 | 1.43* | 0.04 |

[a] Prob{$\pi_{sim} > \pi_{obs}$}, where $\pi_{sim}$ is the expected range of $\pi$ values in the simulation (supplementary fig. S2, Supplementary Material online) using $\theta_A$ values and $\pi_{obs}$ is the observed $\pi$ values.

[b] Prob{$\theta_{wsim} > \theta_{wobs}$}, where $\theta_{wsim}$ is the expected range of $\theta_w$ values in the simulation (supplementary fig. S3, Supplementary Material online) using $\theta_A$ values and $\theta_{wobs}$ is the observed $\theta_w$ values.

[c] Prob{$D_{sim} > D_{obs}$}, where $D_{sim}$ is the expected range of Tajima's $D$ values in the simulation (supplementary fig. S4, Supplementary Material online) using $\theta_A$ values and $D_{obs}$ is the observed Tajima's $D$ values.

*Significant at the 5% level. **Significant at the 1% level.

one of the samples in both strands to confirm the variation. When we found a different nucleotide in only one sample, that is, as a singleton among the samples, we repeated the PCR for the sample and sequenced the region for both strands to rule out a PCR error and confirm the variation. The 28-bp coding exon in the B2m gene and insertion–deletion sites were excluded from the analysis. As was the case with the L-M opsin gene, where we did not do the DNA cloning, we also did not clone the neutral references and thus two allelic DNA sequences of individuals were not separated whose two alleles have different sequences. In the alignment (Supplementary Material online), two heterozygous alleles were tentatively sorted into the allele sequences that were found in homozygous individuals.

## Summary Statistics of Within-Species Nucleotide Variation

Nucleotide diversity ($\pi$) is the average number of nucleotide differences per nucleotide site between two sequences (and is also the unbiased estimator of the average heterozygosity among nucleotide sites) (Nei and Kumar 2000). The number of polymorphic (segregating) sites among samples ($S$) derives a nucleotide polymorphism parameter $\theta_W \equiv S/L/\sum_{i=1}^{n-1} 1/i$, where $L$ is the length of the sequence and $n$ is the number of samples (Watterson 1975). In theory, when mutations are selectively neutral and population size is constant through generations, both $\pi$ and $\theta_w$ are expected to converge to the population mutation rate $\theta$ ($\equiv 4N_e\mu$ for autosomal and $3N_e\mu$ for X-chromosomal genes of diploid organisms, where $N_e$ is the effective population size and $\mu$ is the mutation rate per nucleotide site per chromosome per generation). Tajima's $D$ evaluates the difference between $\pi$ and $\theta_w$, which is given by $\pi - \theta_w$ divided by the estimated standard error of the difference (Tajima 1989). Tajima's $D$ value of neutral references can be regarded as a control measure of severity of demographic effects, such as reduction or expansion of the population size and admixture of genetically differentiated subpopulations, which are considered to affect all genes alike in the genome. On top of it, Tajima's $D$ value of the L-M opsin gene can be regarded as a measure of natural selection operating specifically on it. The summary statistics $\pi$, $\theta_w$, and Tajima's $D$ values were computed using DnaSP version 4.90.1 software (http://www.ub.edu/dnasp/).

## Nucleotide Divergence between Species

The proportion of nucleotide sites fixed with different nucleotides between spider and capuchin monkeys (i.e., divergence between species) was examined using all the samples listed in tables 4 and 5 for the same regions between the two species: for neutral reference, 428 bp of η-globin pseudogene, 457 bp of vWF intron 11, 487 bp of TF intron 5, 226 bp of B2m intron 2, and 447 bp of S opsin intron 4 (total 2,045 bp); for the L-M opsin gene, 881 bp of Exon 1, 949 bp of Exon 3, 827 bp of Exon 5, and 835 bp of Exon 6 (total 3,492 bp).

## Coalescence Simulation

A coalescence simulation was performed to evaluate statistical significance of the deviation of the observed summary statistics in the L-M opsin gene from expectation on the basis of the observed values of summary statistics in the neutral reference loci (Innan 2006). For the coalescence simulation, we used the "sarg" software (Nordborg and Innan 2003), which simulates patterns of nucleotide polymorphism under the infinite site model with recombination. The population size was assumed to be constant through generations. To simulate a pattern of polymorphism when all mutations are neutral, two parameters are required: $\theta$ and the population recombination rate $\rho$ ($\equiv 4N_e r$ for autosomal and $3N_e r$ for X-chromosomal genes of diploid organisms, where $r$ is the recombination rate per nucleotide site per chromosome per generation). In the simulation, $\rho$ was set to be equal to $\theta$ as a first approximation verified for human (Nielsen 2000; Nachman 2001; Pritchard and Przeworski 2001).

First, a coalescent-based rejection-sampling algorithm was applied to estimate $\theta$ from the observed $\pi$ values of the five reference regions (η-globin pseudogene, vWF gene intron 11, TF gene intron 5, B2m gene introns 2 and 3, and S opsin gene intron 4) where all nucleotide variations were assumed to be selectively neutral. The length of the sequence was set to the average of the five reference regions examined, that is, 544 bp for spider monkeys and 500 bp for capuchin monkeys (table 4). The number of sequences was set to twice the total number of individuals examined because the reference regions were all autosomal, that is, 64 for spider monkeys and 46 for capuchin monkeys (table 1). Following Tajima (1989), we created a genealogical relationship for these samples on a random basis and generated the number of mutations in each branch based on randomly assigned value of $\theta$ ($\theta'$) from a range arbitrarily set, 0–7.00 $\times$ $10^{-3}$ for spider monkeys and 0–5.00 $\times$ $10^{-3}$ for capuchin monkeys. To take an error into account in estimating $\theta$ from the five loci, we ran the simulation five times with a $\theta'$ value to simulate patterns of nucleotide variation in five loci, calculated $\pi$ value for each locus as an estimate of $\theta$ and took the average of the five $\pi$ values generated ($\pi_{ave}'$). We accepted the $\theta'$ value if the $\pi_{ave}'$ is within ±1% of the mean of the five $\pi$ values observed (1.75 $\times$ $10^{-3}$ in spider monkeys and 0.80 $\times$ $10^{-3}$ in capuchin monkeys). This process was repeated until 10,000 accepted values of $\theta'$ were accumulated, which represent a sample of the posterior distribution of $\theta$ (supplementary fig. S1, Supplementary Material online) conditional on the observation of $\pi$. The obtained $\theta$ was designated $\theta_A$, meaning $\theta$ of autosomal genes.

Second, another round of coalescence simulation was conducted to test a model of neutrality in the L-M opsin genes. Because $\theta$ is $4N_e\mu$ for autosomal and $3N_e\mu$ for X-chromosomal genes, each of the $\theta_A$ values obtained in the first round simulation was 3/4 folded to obtain $\theta_X$ values representing $\theta$ of X-chromosomal genes. The distribution of $\theta_X$ is given in figure 1. The simulation was performed conditional on the density distribution of $\theta_X$ estimated from the five reference regions, and we obtained the null distributions of summary statistics in the L-M opsin genes. The length of the sequence was set to the total length of the L-M opsin gene examined, that is, 6,041 bp for spider monkeys and 3,958 bp for capuchin monkeys (table 5). The number of sequences was set to the total number of X chromosomes examined, that is, 52 for spider monkeys and 31 for capuchin monkeys (table 1). For each of the 10,000$\theta_X$ values, we created a random genealogical relationship for the samples and generated the number of mutations in each branch by following Tajima (1989) and obtained 10,000 sets of $\pi$ (fig. 2), $\theta_w$ (fig. 3) and Tajima's *D* (fig. 4) values. We then assessed the deviation of the observed $\pi$, $\theta_w$, and Tajima's *D* values of the L-M opsin gene from their expected distribution in the positive direction on the goodness of fit basis. We also carried out this second round simulation by the same procedure but using $\theta_A$ values and evaluated fitting of the observed $\pi$, $\theta_w$, and

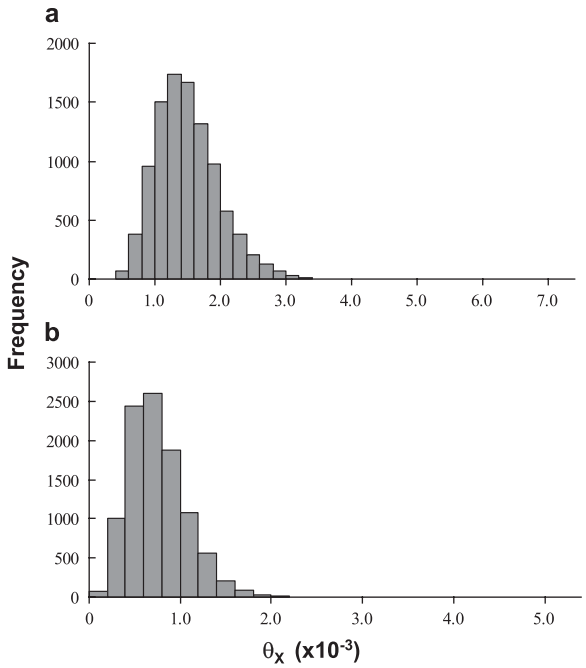


FIG. 1. Samples of the posterior distribution of 10,000$\theta_X$ values in spider monkeys (*a*) and in capuchin monkeys (*b*) obtained by a coalescence simulation conditional on the observation of $\pi$ in the neutral references. Bin size is set to 0.2 $\times$ $10^{-3}$. These distributions essentially represent the likelihood functions of $\theta_X$.

Tajima's *D* values of the neutral references, as well as of the L-M opsin gene regions, to the simulated distributions of these values (table 4, supplementary table S1, supplementary figs. S2, S3, and S4, Supplementary Material online).

## Results

### Color-Vision Phenotypes

In the spider monkey group, the two alleles, P553 and P538, are present at 59.6% and 40.4%, respectively, among 52 X chromosomes examined (table 1). In the capuchin monkey group, the three alleles, P561, P543, and P532, are present at 51.6%, 35.5%, and 12.9%, respectively, among 31 X chromosomes examined. In both species, the genotypes are present at frequencies that do not deviate significantly from the Hardy–Weinberg equilibrium (spider monkeys: $\chi^2 = 0.46$, df = 2, $P > 0.7$; capuchin monkeys: $\chi^2 = 4.39$, df = 5, $P > 0.4$). With the small sample sizes here, it appears that effect of color-vision phenotype on mate choice, if any, is not strong.

### Nucleotide Variation within a Species

The summary statistics of nucleotide variation in the neutral reference regions and in the L-M opsin-gene regions are listed in tables 4 and 5, respectively. Overall, the observed level of polymorphism (i.e., $\pi$ and $\theta_w$) of the L-M opsin gene regions is higher than those of the neutral reference regions in spider and capuchin monkeys. Tajima's *D* value is also higher overall in the L-M opsin gene regions than in the neutral references. Within the L-M opsin gene, the

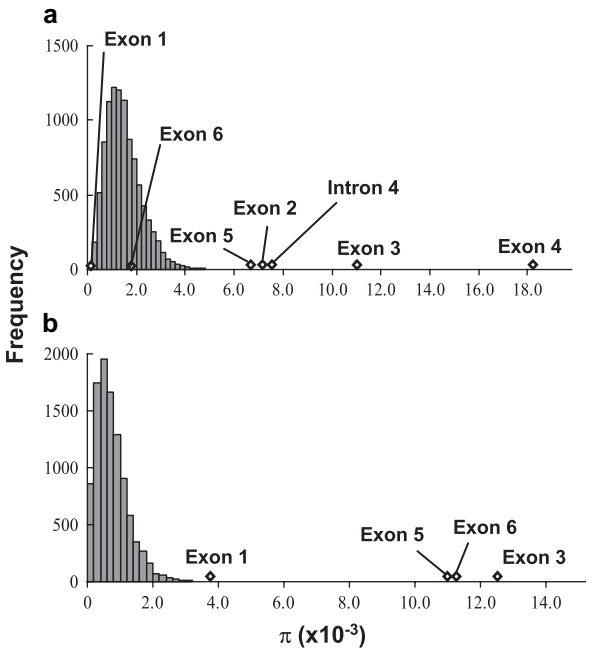


FIG. 2. Expected distribution of 10,000$\pi$ values in the sample size of L-M opsin gene of spider monkeys (*a*) and capuchin monkeys (*b*) obtained by a coalescence simulation based on the distribution of $\theta_X$ shown in figure 1. Bin size is set to $0.2 \times 10^{-3}$. The observed $\pi$ values of the L-M opsin-gene regions are plotted.

peripheral regions (Exons 1 and 6) tend to be less polymorphic than central regions including Exons 3 and 5 where amino acid sites critical for spectral tuning are encoded (table 5).

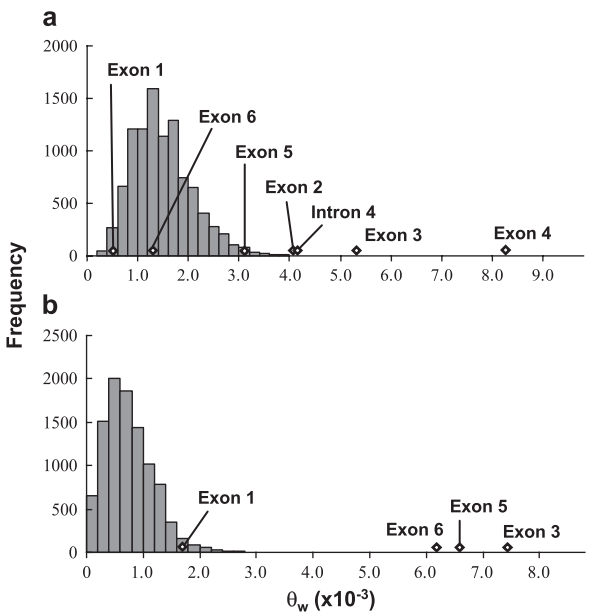


FIG. 3. Expected distribution of 10,000$\theta_w$ values in the sample size of L-M opsin gene of spider monkeys (*a*) and capuchin monkeys (*b*) obtained by a coalescence simulation based on the distribution of $\theta_X$ shown in figure 1. Bin size is set to $0.2 \times 10^{-3}$. The observed $\theta_w$ values of the L-M opsin-gene regions are plotted.

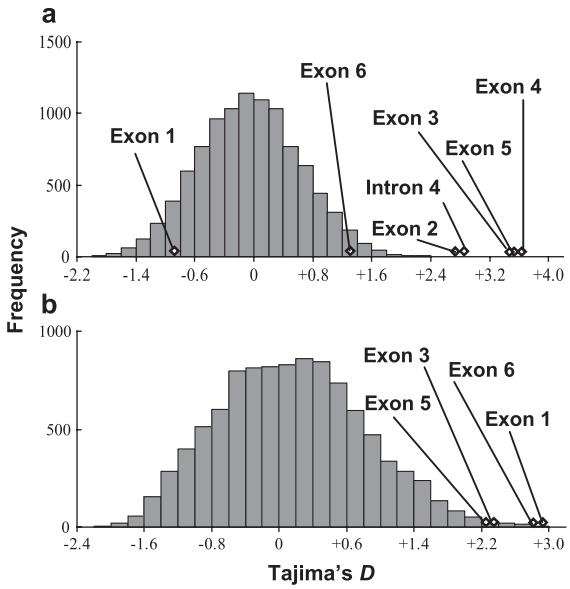


FIG. 4. Expected distribution of 10,000 Tajima's *D* values in the sample size of L-M opsin gene of spider monkeys (*a*) and capuchin monkeys (*b*) obtained by a coalescence simulation based on the distribution of $\theta_X$ shown in figure 1. Bin size is set to $0.2 \times 10^{-3}$. The observed Tajima's *D* values of the L-M opsin-gene regions are plotted.

In capuchin monkeys, there was no nucleotide variation within any of the spectral alleles, P562, P543, and P531, in any of the four regions examined, Exon 1, Exon 3, Exon 5, and Exon 6. Considering that absorption spectra of primate L-M opsins are largely determined by the residues 180 in exon 3 and 277 and 285 in exon 5 (Yokoyama and Radlwimmer 2001; Hiramatsu et al. 2004), it is unlikely that other spectrally different alleles are present in the capuchin monkey group. In spider monkeys, there was no amino acid variation within either of the spectral alleles, P553 and P538, throughout the coding regions. This guarantees that there are no other spectrally different alleles in the spider monkey group. However, there were within-allele nucleotide variations in spider monkeys. The nucleotide diversity ($\pi$) within a spectral allele was calculated for each of the seven regions (table 6). Overall, nucleotide diversity within a spectral allele was much lower than the total nucleotide diversity calculated from all samples except in Exon 1. When we excluded the heterozygous females from the calculation, because their two allele sequences were only tentatively separated (see Materials and Methods), essentially the same pattern, much lower nucleotide diversity within a spectral allele than the total nucleotide diversity, was observed. Thus, in both species, the extent of nucleotide variation within each spectral allele is low and the high degree of polymorphism of the L-M opsin gene regions results from high differentiation between the spectral alleles.

## Nucleotide Divergence between Species

Conversely, the proportion of nucleotide sites fixed differently between spider and capuchin monkeys (i.e., divergence between the two species) is not lower in the neutral references

**Table 5.** Summary Statistics of Nucleotide Variation in the L-M Opsin-Gene Regions of Spider and Capuchin Monkeys.

| Species | Region | No. of Chromosomes Examined ($n$) | No. of Segregating Sites ($S$) | $\pi$ ($\times 10^{-3}$) | $P^a$ | $\theta_w$ ($\times 10^{-3}$) | $P^b$ | Tajima's $D$ | $P^c$ |
|---|---|---|---|---|---|---|---|---|---|
| Spider monkeys | Exon 1 (940 bp) | 48 | 2 | 0.17 | 1 | 0.48 | 1 | −1.16 | 0.9 |
| | Exon 2 (1,015 bp) | 44 | 18 | 7.38** | <0.0001 | 4.08** | 0.0009 | 2.60** | 0.0001 |
| | Exon 3 (1,021 bp) | 46 | 24 | 11.00** | <0.0001 | 5.35** | <0.0001 | 3.51** | <0.0001 |
| | Exon 4 (465 bp) | 48 | 17 | 18.00** | <0.0001 | 8.24** | <0.0001 | 3.73** | <0.0001 |
| | Intron 4 (827 bp) | 48 | 15 | 7.72** | <0.0001 | 4.09** | 0.0009 | 2.75** | <0.0001 |
| | Exon 5 (874 bp) | 52 | 12 | 6.69** | <0.0001 | 3.04* | 0.02 | 3.53** | <0.0001 |
| | Exon 6 (899 bp) | 46 | 5 | 1.92 | 0.2 | 1.27 | 0.6 | 1.28* | 0.03 |
| | Total (6,041 bp) | 43 | 93 | 6.82** | <0.0001 | 3.52** | 0.004 | 3.39** | <0.0001 |
| Capuchin monkeys | Exon 1 (1,026 bp) | 31 | 8 | 3.86** | 0.0003 | 1.95* | 0.02 | 2.94** | 0.0001 |
| | Exon 3 (1,072 bp) | 31 | 32 | 12.49** | <0.0001 | 7.47** | <0.0001 | 2.42** | 0.004 |
| | Exon 5 (1,024 bp) | 31 | 27 | 11.06** | <0.0001 | 6.60** | <0.0001 | 2.41** | 0.004 |
| | Exon 6 (836 bp) | 31 | 21 | 11.30** | <0.0001 | 6.29** | <0.0001 | 2.77** | 0.001 |
| | Total (3,958 bp) | 31 | 88 | 9.59** | <0.0001 | 5.57** | <0.0001 | 2.74** | 0.001 |

[a] Prob{$\pi_{sim} > \pi_{obs}$}, where $\pi_{sim}$ is the expected range of $\pi$ values in the simulation (fig. 2) using $\theta_X$ values and $\pi_{obs}$ is the observed $\pi$ values.
[b] Prob{$\theta_{wsim} > \theta_{wobs}$}, where $\theta_{wsim}$ is the expected range of $\theta_w$ values in the simulation (fig. 3) using $\theta_X$ values and $\theta_{wobs}$ is the observed $\theta_w$ values.
[c] Prob{$D_{sim} > D_{obs}$}, where $D_{sim}$ is the expected range of Tajima's $D$ values in the simulation (fig. 4) using $\theta_X$ values and $D_{obs}$ is the observed Tajima's $D$ values.
*Significant at the 5% level. **Significant at the 1% level.

(5.0%; 103/2,045) than in the L-M opsin gene region (3.2%; 113/3,492) (table 7). This indicates that the mutation rate is not higher in the L-M opsin gene than in the neutral reference regions and does not explain the higher within-species variation in the L-M opsin gene. Within the L-M opsin gene, similarly, the divergence between spider and capuchin monkeys is not lower in the less polymorphic peripheral Exon 1 (4.0%; 35/881) and Exon 6 (4.7%; 39/835) regions than in the central Exon 3 (2.7%; 26/949) and Exon 5 (1.6%; 13/827) regions (table 7). This indicates that the mutation rate is also not higher in the central than in the peripheral regions and does not explain the higher within-species variation in the central region.

**Table 6.** Nucleotide Variation within a Spectral Allele of the L-M Opsin Gene in Spider Monkeys.

| Region | No. of Chromosomes Examined ($n$) | Length Analyzed ($L$) (bp) | No. of Segregating Sites ($S$) | $\pi$ ($\times 10^{-3}$) |
|---|---|---|---|---|
| Exon 1 | 48 | 940 | 2 | 0.17 |
| P553 | 28 | 940 | 1 | 0.21 |
| P538 | 20 | 940 | 1 | 0.11 |
| Exon 2 | 44 | 1,015 | 18 | 7.38 |
| P553 | 26 | 1,015 | 4 | 0.89 |
| P538 | 18 | 1,015 | 8 | 1.31 |
| Exon 3 | 46 | 1,021 | 24 | 11.04 |
| P553 | 27 | 1,021 | 7 | 3.13 |
| P538 | 19 | 1,021 | 3 | 0.49 |
| Exon 4 | 48 | 465 | 17 | 18.02 |
| P553 | 27 | 465 | 1 | 1.04 |
| P538 | 21 | 465 | 0 | 0.00 |
| Intron 4 | 48 | 827 | 15 | 7.72 |
| P553 | 28 | 827 | 8 | 3.76 |
| P538 | 20 | 827 | 2 | 0.60 |
| Exon 5 | 52 | 874 | 12 | 6.69 |
| P553 | 31 | 874 | 1 | 0.59 |
| P538 | 21 | 874 | 0 | 0.00 |
| Exon 6 | 46 | 899 | 5 | 1.92 |
| P553 | 27 | 899 | 1 | 0.08 |
| P538 | 19 | 899 | 1 | 0.46 |
| Total | 43 | 6,041 | 93 | 6.82 |
| P553 | 25 | 6,041 | 23 | 1.43 |
| P538 | 18 | 6,041 | 15 | 0.43 |

## Coalescence Simulation

The coalescence simulation showed that in both spider and capuchin monkeys the observed $\pi$ values of the L-M opsin gene regions were apparent outliers in the expected distribution except for Exons 1 and 6 in spider monkeys (fig. 2; see also table 5 for $P$ values) whereas, as expected, those of the neutral reference regions were within the distribution (supplementary fig.S2, Supplementary Material online; see table 4 for $P$ values). These results indicate that the observed $\pi$ values of the L-M opsin gene are significantly larger than those of the neutral reference regions in both spider and capuchin monkeys. A similar pattern was observed for distribution of $\theta_w$ values, that is, $\theta_w$ values of the neutral reference regions are within the expected range (supplementary fig. S3, Supplementary Material online; see table 4 for $P$ values), whereas those of the L-M opsin-gene regions tend to be outliers in a positive direction (fig. 3; see table 5 for $P$ values). However, the pattern is less conspicuous than that in $\pi$ values. The distribution of the neutral reference regions is biased toward the smaller values in the expected distribution, and that of the L-M opsin gene Exons 1 and 6 in spider monkeys is within the expected range.

**Table 7.** Nucleotide Divergence between Spider and Capuchin Monkeys.

| Region | Length (bp) | No. of Fixed Sites | Divergence (%) |
|---|---|---|---|
| Neutral Reference | 2,045 | 103 | 5.0 |
| L-M opsin | | | |
| Exon 1 | 881 | 35 | 4.0 |
| Exon 3 | 949 | 26 | 2.7 |
| Exon 5 | 827 | 13 | 1.6 |
| Exon 6 | 835 | 39 | 4.7 |
| Total | 3,492 | 113 | 3.2 |

Regarding the neutral reference regions, the difference of $\pi$ and $\theta_w$ in the distribution of observed values in the expected range resulted in positive values of Tajima's *D*, especially in capuchin monkeys (supplementary fig. S4, Supplementary Material online; see table 4 for *P* values). This might suggest a degree of long-term reduction in population size or of population admixture in both species. However, with reference to Tajima's *D* values of the L-*M* opsin-gene regions, the deviation of observation from the expected range was far more conspicuous and significant in the positive direction, except in Exon 1 in spider monkeys (fig. 4; see table 5 for *P* values).

Recent studies suggest that $\theta$ of X-chromosomal genes is not exactly 3/4-fold of that of autosomal genes (Sachidanandam et al. 2001; Hammer et al. 2008; Keinan et al. 2009). To be conservative, when we assumed $\theta_X/\theta_A = 1$ for the coalescence simulation, most of the observed $\pi$, $\theta_w$ and Tajima's *D* values of the X-chromosomal L-*M* opsin-gene regions still deviated significantly from the expected range (supplementary figs. S2, S3, and S4, Supplementary Material online; see supplementary table S1, Supplementary Material online, for *P* values). With recombination assumed in the coalescence simulation, the distribution of $\pi$, $\theta_w$ and Tajima's *D* values is expected to be narrowed while leaving their mean values unaffected. When we set the population recombination rate $\rho$ to zero in the simulation using $\theta_A$ values, which would give the broadest distribution of these values, the deviation of the observed values from the expected values was still evident (supplementary table S1, Supplementary Material online).

Based on the coalescence simulation with the neutral reference regions as the controls, we conclude that balancing selection is operating on the centrally located exons of the L-*M* opsin gene, which encode the amino acid sites critical for spectral tuning in both of these two distantly related species of New World monkeys.

## Discussion

In this study, we detected an explicit signal of balancing selection for color-vision variation in New World monkeys by examining the nucleotide variation of the L-*M* opsin gene region and, as a neutral reference, other genome regions for natural populations of spider monkeys and capuchin monkeys. We took a computer-simulation approach to statistically test the deviation of the extent of nucleotide variation in the L-*M* opsin gene from the simulated values based on the observed extent of nucleotide variation in the neutral reference regions. This approach is robust against the effect of the demography of the populations studied such as population size change, migration, and population structure. The demography affects the nucleotide variation of all the genomic regions in a similar way. Because our computer simulation approach focuses on the difference of the levels of nucleotide variation of the L-*M* opsin gene and the reference regions, the demographic effect for the L-*M* opsin gene and the reference regions is canceled out. In this study, the standard Tajima's *D* test was also highly significant for the L-*M* opsin gene (table 5). This result also suggests that the positive natural selection operated on the L-*M* gene. However, the standard Tajima's *D* test is sensitive to the demographic effect. The possibility that Tajima's *D* became significantly positive for the L-*M* opsin gene due to the population size reduction or population structure could not be excluded if the nucleotide variation of the reference regions were not compared.

We noted that the high nucleotide variation in the L-*M* opsin gene was due to the difference between the allele types and that there was only low (in spider monkeys: table 6) or no (in capuchin monkeys) variation within a given allele type. This suggests that the populations are not large enough to hold nucleotide variation within an allele type under the effect of random genetic drift, which in turn emphasizes the large magnitude of balancing selection to maintain the difference between alleles. It should be noted that the population size of the capuchin monkeys would be smaller than that of the spider monkeys because $\pi$ and $\theta_w$ values of neutral reference genes are smaller for capuchin monkeys than for spider monkeys (table 4).

A measure of linkage disequilibrium (LD), $D'$, which takes a range from 0 to 1 (Hartl and Clark 2007), was calculated for the L-*M* opsin gene regions using male and homozygous female samples. The $D'$ was nearly 1 throughout the gene regions in both spider monkeys and capuchin monkeys (data not shown). This indicates low incidence of recombination between alleles. Interspecies comparative studies have shown that phylogenetic trees constructed using the intron 4 sequences of the L-*M* opsin gene from New World monkeys show closer relatedness between alleles within a species than between the same spectral alleles of different species, whereas those using exons 3, 4, and 5 show an opposite pattern which is consistent with ancient origin of these alleles (Shyue et al. 1995; Boissinot et al. 1998). These studies argued for interallelic recombination (designated "gene conversion" in the literatures) in intron 4 and the balancing selection against it on the exons. LD is affected by population size and structure and reflects relatively recent incidence of recombination. The low incidence of recombination shown by high LD is not incompatible with the suggested recombination distorting the phylogenetic tree during long evolutionary time frame. The low intraspecies nucleotide variation in the peripheral regions of the L-*M* opsin gene found in this study, in spite of the high variation in the central region by balancing selection and the similar interspecies divergence between the two regions, can also be explained by the low but long-term incidence of recombination between alleles. This could effectively separate the peripheral regions from the influence of balancing selection on the central region. Yet, the suggested recombination in intron 4 (Shyue et al. 1995; Boissinot et al. 1998), located centrally between exon 4 and exon 5, is not evident in this study showing high intraspecies variation (table 5). This suggests that the incidence of recombination in the central region not involving exons is even lower and is not effective enough

to result in lowering genetic variation, though it is effective in distorting the phylogenetic tree.

Balancing selection has often been invoked on the basis of unequal allele frequencies of the L-M opsin gene among samples of New World monkeys (Osorio et al. 2004; Surridge, Suarez, Buchanan-Smith, Smith, and Mundy 2005). In our study site, allele frequency of the L-M opsin gene was higher for the longer wave sensitive alleles in both spider and capuchin monkeys (table 1). Neutral alleles can appear at unequal frequencies (Hartl and Clark 2007), and this inequality itself cannot be a sign of natural selection. The allele-frequency composition is variable among species studied (Cropp et al. 2002; Surridge and Mundy 2002; Surridge, Suarez, Buchanan-Smith, Smith, and Mundy 2005), and more population data would be necessary to determine the nature of selection based on the pattern of allele-frequency composition.

Given the clear indication of balancing selection obtained in our study and the uncertainty about benefits of trichromacy (i.e., for individuals heterozygous for the L-M opsin alleles) obtained from behavioral studies of wild primates (Dominy et al. 2003; Smith, Buchanan-Smith, Surridge, and Mundy 2003; Melin et al. 2007; Vogel et al. 2007; Hiramatsu et al. 2008, 2009), how should we interpret the nature of balancing selection? The advantages of trichromacy might be manifested in tasks for which behavioral data have yet to be gathered, such as long-distance detection of reddish objects under dappled foliage (Sumner and Mollon 2000), foraging on reddish ripe fruits in severe dry seasons when these could be scarce (Dominy and Lucas 2001), or recognition of social signals (Changizi et al. 2006; Fernandez and Morris 2007). Besides the trichromat advantage, three other mechanisms of balancing selection have been hypothesized to explain color-vision polymorphism in New World monkeys (Mollon et al. 1984): 1) negative frequency–dependent selection, which predicts the fitness of any given phenotype to be reciprocal to the frequency of that phenotype in the population; 2) niche divergence, which predicts that individuals of each phenotype will specialize in a distinct visual or ecological niche or visual ability; and 3) mutual benefit of association, which predicts that individuals of each phenotype benefit from being associated with individuals of other phenotypes in a polymorphic group.

Negative frequency–dependent selection is generally invoked for predator–prey interaction or a disassortative mating system, that is, mating with a different type from oneself in terms of the genetic trait in question (Conner and Hartl 2004). This is hard to envision in the case of color vision and often appears to be interpreted mistakenly in the literature as a consequence of niche divergence (thus the two hypotheses are often confounded). Under the niche-divergence situation, the population size of each phenotype can fluctuate independently of each other and irrespectively of its frequency because individuals with different phenotypes exploit different resources or niches, and population size of the phenotype changes as the carrying capacity of their niche changes but not as population size of another phenotype changes. Although negative frequency–dependent selection is often referred to as an alternative explanation to the heterozygote advantage hypothesis as a mechanism of maintaining color-vision polymorphism, it would be the least likely mechanism. One recent field study reports disassortative mating with respect to color-vision genotype among tamarin monkeys, but it reports that this is likely driven by a general inbreeding avoidance rather than any mechanism specific to the opsin locus (Surridge, Suarez, Buchanen-Smith, and Mundy, 2005). Although inbreeding certainly reduces genetic variability in a general sense, simple avoidance of it, and otherwise random mating, cannot stop the reduction of genetic variability in successive generations in finite populations. The niche-divergence hypothesis has also received no support from field observations: In the food foraging behaviors of capuchin monkeys, there is no difference in foraging time spent on different food types between dichromatic and trichromatic individuals (Melin et al. 2008).

Although no study has evaluated the mutual benefit hypothesis, balancing selection may represent an advantage of individuals with different color-vision phenotypes coexisting in the same population. Monkey and ape dichromats, as well as human dichromats, exhibit a clear advantage in detecting color-camouflaged objects (Morgan et al. 1992; Caine et al. 2003; Saito, Mikami, et al. 2005), including surface-dwelling insects (Melin et al. 2007), an important food source for many primates. Humans are polymorphic in color vision and have a long history of a hunting and gathering lifestyle in which the ability to break camouflage may be advantageous. Interdisciplinary studies, like ours, combining genetics, behavioral ecology, and visual physiology, provide a wealth of data for furthering our understanding of the evolution of color vision.

## Supplementary Material

Supplementary figures S1–S4, supplementary table S1, and all of the sequence alignments used in this study are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

## Acknowledgments

We thank R. Blanco Segura, M. M. Chavarria, and other staff of the Area de Conservación Guanacaste for local support and are grateful to the Ministerio de Ambiente y Energía (MINAE) of Costa Rica for giving us permission to conduct this study at Santa Rosa National Park. We appreciate E. Murillo Chacon, C. Sendall, K. M. Jack, S. Carnegie, and L. Rebecchini for their assistance in collecting fecal samples. The manuscript has been improved thanks to comments of Naoko Takezaki and anonymous reviewers. This study was supported by Grants-in-Aid for Scientific Research (B) (16405015) and (A) (19207018) from the Japan Society for the Promotion of Science to S.K.

## References


Aureli F, Schaffner CM. 2008. Spider monkeys: social structure, social relationships and social interactions. In: Campbell C, editor. Spider monkeys: behavior ecology & evolution of the genus *Ateles*. Cambridge (MA): Cambridge University Press. p. 236–265.

Bailey WJ, Fitch DH, Tagle DA, Czelusniak J, Slightom JL, Goodman M. 1991. Molecular evolution of the psi eta-globin gene locus: gibbon phylogeny and the hominoid slowdown. *Mol Biol Evol*. 8:155–184.

Boissinot S, Tan Y, Shyue SK, Schneider H, Sampaio I, Neiswanger K, Hewett-Emmett D, Li WH. 1998. Origins and antiquity of X-linked triallelic color vision systems in New World monkeys. *Proc Natl Acad Sci USA*. 95:13749–13754.

Caine NG. 2002. Seeing red: consequences of individual differences in color vision in callitrichid primates. In: Miller LE, editor. Eat or be eaten. Cambridge (MA): Cambridge University Press. p. 58–73

Caine NG, Mundy NI. 2000. Demonstration of a foraging advantage for trichromatic marmosets (*Callithrix geoffroyi*) dependent on food colour. *Proc R Soc Lond B*. 267:439–444.

Caine NG, Surridge AK, Mundy NI. 2003. Dichromatic and trichromatic *Callithrix geoffroyi* differ in relative foraging ability for red-green color-camouflaged and non-camouflaged food. *Int J Primatol*. 24:1163–1175.

Canavez FC, Ladasky JJ, Muniz JA, Seuanez HN, Parham P, Cavanez C. 1998. beta2-Microglobulin in neotropical primates (Platyrrhini). *Immunogenetics* 48:133–140.

Changizi MA, Zhang Q, Shimojo S. 2006. Bare skin, blood and the evolution of primate color vision. *Biol Lett*. 2:217–221.

Conner JK, Hartl DL. 2004. A primer of ecological genetics. Sunderland (MA): Sinauer Associates.

Cropp S, Boinski S, Li WH. 2002. Allelic variation in the squirrel monkey X-linked color vision gene: biogeographical and behavioral correlates. *J Mol Evol*. 54:734–745.

De Araujo MF, Lima EM, Pessoa VF. 2006. Modeling dichromatic and trichromatic sensitivity to the color properties of fruits eaten by squirrel monkeys (*Saimiri sciureus*). *Am J Primatol*. 68:1129–1137.

Dominy NJ, Garber PA, Bicca-Marques JC, Azevedo-Lopes MA. 2003. Do female tamarins use visual cues to detect fruit rewards more successfully than do males? *Anim Behav*. 66:829–837.

Dominy NJ, Lucas PW. 2001. Ecological importance of trichromatic vision to primates. *Nature* 410:363–366.

Dominy NJ, Lucas PW. 2004. Significance of color, calories, and climate to the visual ecology of catarrhines. *Am J Primatol*. 62:189–207.

Fernandez AA, Morris MR. 2007. Sexual selection and trichromatic color vision in primates: statistical support for the preexisting-bias hypothesis. *Am Nat*. 170:10–20.

Fitch DH, Mainone C, Slightom JL, Goodman M. 1988. The spider monkey psi eta-globin gene and surrounding sequences: recent or ancient insertions of LINEs and SINEs? *Genomics* 3:237–255.

Fragaszy DM, Visalberghi E, Fedigan LM. 2004. The complete capuchin: the biology of the genus *Cebus*. Cambridge (MA): Cambridge University Press.

Hammer MF, Mendez FL, Cox MP, Woerner AE, Wall JD. 2008. Sex-biased evolutionary forces shape genomic patterns of human diversity. *PLoS Genet*. 4:e1000202.

Hartl DL, Clark AG. 2007. Principles of population genetics. 4th ed. Sunderland (MA): Sinauer Associates.

Hiramatsu C, Melin AD, Aureli F, Schaffner CM, Vorobyev M, Kawamura S. 2009. Interplay of olfaction and vision in fruit foraging of spider monkeys. *Anim Behav*. 77:1421–1426.

Hiramatsu C, Melin AD, Aureli F, Schaffner CM, Vorobyev M, Matsumoto Y, Kawamura S. 2008. Importance of achromatic contrast in short-range fruit foraging of primates. *PLoS ONE*. 3:e3356.

Hiramatsu C, Radlwimmer FB, Yokoyama S, Kawamura S. 2004. Mutagenesis and reconstitution of middle-to-long-wave-sensitive visual pigments of New World monkeys for testing the tuning effect of residues at sites 229 and 233. *Vision Res*. 44:2225–2231.

Hiramatsu C, Tsutsui T, Matsumoto Y, Aureli F, Fedigan LM, Kawamura S. 2005. Color-vision polymorphism in wild capuchins (*Cebus capucinus*) and spider monkeys (*Ateles geoffroyi*) in Costa Rica. *Am J Primatol*. 67:447–461.

Innan H. 2006. Modified Hudson–Kreitman–Aduae test and two-dimensional evaluation of neutrality tests. *Genetics* 173: 1725–1733.

Jacobs GH. 1996. Primate photopigments and primate color vision. *Proc Natl Acad Sci USA*. 93:577–581.

Jacobs GH. 2007. New World monkeys and color. *Int J Primatol*. 28:729–759.

Kawamura S, Hirai M, Takenaka O, Radlwimmer FB, Yokoyama S. 2001. Genomic and spectral analyses of long to middle wavelength-sensitive visual pigments of common marmoset (*Callithrix jacchus*). *Gene*. 269:45–51.

Keinan A, Mullikin JC, Patterson N, Reich D. 2009. Accelerated genetic drift on chromosome X during the human dispersal out of Africa. *Nat Genet*. 41:66–70.

Koop BF, Goodman M, Xu P, Chan K, Slightom JL. 1986. Primate eta-globin DNA sequences and man's place among the great apes. *Nature* 319:234–238.

Lucas PW, Darvell BW, Lee PKD, Yuen TDB, Choong MF. 1998. Colour cues for leaf food selection by long-tailed macaques (*Macaca fascicularis*) with a new suggestion for the evolution of trichromatic colour vision. *Folia Primatol*. 69:139–154.

Lucas PW, Dominy NJ, Riba-Hernandez P, et al. (12 co-authors). 2003. Evolution and function of routine trichromatic vision in primates. *Evolution* 57:2636–2643.

Melin AD, Fedigan LM, Hiramatsu C, Kawamura S. 2008. Polymorphic color vision in white-faced capuchins (*Cebus capucinus*): is there foraging niche divergence among phenotypes? *Behav Ecol Sociobiol*. 62:659–670.

Melin AD, Fedigan LM, Hiramatsu C, Sendall C, Kawamura S. 2007. Effects of colour vision phenotype on insect capture by a free-ranging population of white-faced capuchins (*Cebus capucinus*). *Anim Behav*. 73:205–214.

Mollon JD, Bowmaker JK, Jacobs GH. 1984. Variations of colour vision in a New World primate can be explained by polymorphism of retinal photopigments. *Proc R Soc Lond B*. 222:373–399.

Morgan MJ, Adam A, Mollon JD. 1992. Dichromats detect colour-camouflaged objects that are not detected by trichromats. *Proc R Soc Lond B*. 248:291–295.

Nachman MW. 2001. Single nucleotide polymorphisms and recombination rate in humans. *Trends Genet*. 17:481–485.

Nagao K, Takenaka N, Hirai M, Kawamura S. 2005. Coupling and decoupling of evolutionary mode between X- and Y-chromosomal red-green opsin genes in owl monkeys. *Gene*. 352:82–91.

Nathans J, Thomas D, Hogness DS. 1986. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* 232:193–202.

Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. New York: Oxford university press.

Neitz M, Neitz J, Jacobs GH. 1991. Spectral tuning of pigments underlying red-green color vision. *Science*. 252:971–974.

Nielsen R. 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 154:931–942.

Nordborg M, Innan H. 2003. The genealogy of sequences containing multiple sites subject to strong selection in a subdivided population. *Genetics* 163:1201–1213.

Osorio D, Smith AC, Vorobyev M, Buchanan-Smith HM. 2004. Detection of fruit and the selection of primate visual pigments for color vision. *Am Nat*. 164:696–708.

Parraga CA, Troscianko T, Tolhurst DJ. 2002. Spatiochromatic properties of natural images and human vision. *Curr Biol*. 12:483–487.

Perry GH, Martin RD, Verrelli BC. 2007. Signatures of functional constraint at aye-aye opsin genes: the potential of adaptive color vision in a nocturnal primate. *Mol Biol Evol*. 24:1963–1970.

Pritchard JK, Przeworski M. 2001. Linkage disequilibrium in humans: models and data. *Am J Hum Genet*. 69:1–14.

Regan BC, Julliot C, Simmen B, Vienot F, Charles-Dominique P, Mollon JD. 1998. Frugivory and colour vision in *Alouatta seniculus*, a trichromatic platyrrhine monkey. *Vision Res*. 38:3321–3327.

Regan BC, Julliot C, Simmen B, Vienot F, Charles-Dominique P, Mollon JD. 2001. Fruits, foliage and the evolution of primate colour vision. *Phil Trans R Soc Lond B*. 356:229–283.

Riba-Hernandez P, Stoner KE, Lucas PW. 2005. Sugar concentration of fruits and their detection via color in the Central American spider monkey (*Ateles geoffroyi*). *Am J Primatol*. 67:411–423.

Riba-Hernandez P, Stoner KE, Osorio D. 2004. Effect of polymorphic colour vision for fruit detection in the spider monkey *Ateles geoffroyi*, and its implications for the maintenance of polymorphic colour vision in platyrrhine monkeys. *J Exp Biol*. 207:2465–2470.

Sachidanandam R, Weissman D, Schmidt SC, et al. (41 co-authors). 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409: 928–933.

Saito A, Kawamura S, Mikami A, Ueno Y, Hiramatsu C, Koida K, Fujita K, Kuroshima H, Hasegawa T. 2005. Demonstration of a genotype-phenotype correlation in the polymorphic color vision of a non-callitrichine New World monkey, capuchin (*Cebus apella*). *Am J Primatol*. 67:471–485.

Saito A, Mikami A, Kawamura S, et al. (13 co-authors). 2005. Advantage of dichromats over trichromats in discrimination of color-camouflaged stimuli in nonhuman primates. *Am J Primatol*. 67:425–436.

Schneider H. 2000. The current status of the New World monkey phylogeny. *An Acad Bras Ci*. 72:165–172.

Shimmin LC, Miller J, Tran HN, Li WH. 1998. Contrasting levels of DNA polymorphism at the autosomal and X-linked visual color pigment loci in humans and squirrel monkeys. *Mol Biol Evol*. 15:449–455.

Shyue SK, Boissinot S, Schneider H, et al. (13 co-authors). 1998. Molecular genetics of spectral tuning in New World monkey color vision. *J Mol Evol*. 46:697–702.

Shyue SK, Hewett-Emmett D, Sperling HG, Hunt DM, Bowmaker JK, Mollon JD, Li WH. 1995. Adaptive evolution of color vision genes in higher primates. *Science* 269:1265–1267.

Smith AC, Buchanan-Smith HM, Surridge AK, Mundy NI. 2003. Leaders of progressions in wild mixed-species troops of saddleback (*Saguinus fuscicollis*) and mustached tamarins (*S. mystax*), with emphasis on color vision and sex. *Am J Primatol*. 61:145–157.

Smith AC, Buchanan-Smith HM, Surridge AK, Osorio D, Mundy NI. 2003. The effect of colour vision status on the detection and selection of fruits by tamarins (*Saguinus* spp.). *J Exp Biol*. 206:3159–3165.

Stoner KE, Riba-Hernandez P, Lucas PW. 2005. Comparative use of color vision for frugivory by sympatric species of platyrrhines. *Am J Primatol*. 67:399–409.

Sumner P, Mollon JD. 2000. Catarrhine photopigments are optimized for detecting targets against a foliage background. *J Exp Biol*. 203:1963–1986.

Sumner P, Mollon JD. 2003. Colors of primate pelage and skin: objective assessment of conspicuousness. *Am J Primatol*. 59:67–91.

Sun H, Macke JP, Nathans J. 1997. Mechanisms of spectral tuning in the mouse green cone pigment. *Proc Natl Acad Sci USA*. 94:8860–8865.

Surridge AK, Mundy NI. 2002. Trans-specific evolution of opsin alleles and the maintenance of trichromatic colour vision in Callitrichine primates. *Mol Ecol*. 11:2157–2169.

Surridge AK, Osorio D, Mundy NI. 2003. Evolution and selection of trichromatic vision in primates. *Trends Ecol Evol*. 18:198–205.

Surridge AK, Suarez SS, Buchanan-Smith HM, Mundy NI. 2005. Non-random association of opsin alleles in wild groups of red-bellied tamarins (*Saguinus labiatus*) and maintenance of the colour vision polymorphism. *Biol Lett*. 1:465–468.

Surridge AK, Suarez SS, Buchanan-Smith HM, Smith AC, Mundy NI. 2005. Color vision pigment frequencies in wild tamarins (*Saguinus* spp.). *Am J Primatol*. 67:463–470.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Verrelli BC, Lewis CM Jr., Stone AC, Perry GH. 2008. Different selective pressures shape the molecular evolution of color vision in chimpanzee and human populations. *Mol Biol Evol*. 25:2735–2743.

Verrelli BC, Tishkoff SA. 2004. Signatures of selection and gene conversion associated with human color vision variation. *Am J Hum Genet*. 75:363–375.

Vogel ER, Neitz M, Dominy NJ. 2007. Effect of color vision phenotype on the foraging of wild white-faced capuchins, *Cebus capucinus*. *Behav Ecol*. 18:292–297.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 7:256–276.

Yokoyama R, Yokoyama S. 1990. Convergent evolution of the red- and green-like visual pigment genes in fish, *Astyanax fasciatus*, and human. *Proc Natl Acad Sci USA*. 87:9315–9318.

Yokoyama S, Radlwimmer FB. 1998. The "five-sites" rule and the evolution of red and green color vision in mammals. *Mol Biol Evol*. 15:560–567.

Yokoyama S, Radlwimmer FB. 2001. The molecular genetics and evolution of red and green color vision in vertebrates. *Genetics* 158:1697–1710.

Yokoyama S, Yang H, Starmer WT. 2008. Molecular basis of spectral tuning in the red- and green-sensitive (M/LWS) pigments in vertebrates. *Genetics* 179:2037–2043.

**Table S1**

**Summary statistics of nucleotide variation in the L-M opsin gene regions of spider and capuchin monkeys**

| Species | Region | No. of chromosomes examined (*n*) | No. of segregating sites (*S*) | $\pi$ (x $10^{-3}$) | *p* [a] | $\theta_w$ (x $10^{-3}$) | *p* [b] | Tajima's *D* | *p* [c] |
|---|---|---|---|---|---|---|---|---|---|
| Spider monkeys | Exon 1 (940 bp) | 48 | 2 | 0.17 | 1 (1) | 0.48 | 1 (1) | -1.16 | 1 (1) |
| | Exon 2 (1015 bp) | 44 | 18 | 7.38** | 0.0001 (0.004) | 4.08** | 0.001 (0.003) | 2.60** | 0.0001 (<0.0001) |
| | Exon 3 (1021 bp) | 46 | 24 | 11.00** | <0.0001 (<0.001) | 5.35** | 0.0008 (0.004) | 3.51** | <0.0001 (<0.0001) |
| | Exon 4 (465 bp) | 48 | 17 | 18.00** | <0.0001 (<0.0001) | 8.24** | <0.0001 (0.0002) | 3.73** | <0.0001 (<0.0001) |
| | Intron 4 (827 bp) | 48 | 15 | 7.72** | <0.0001 (0.003) | 4.09** | 0.001 (0.003) | 2.75** | 0.0001 (<0.0001) |
| | Exon 5 (874 bp) | 52 | 12 | 6.69** | 0.0001 (0.007) | 3.04 | 0.1 (0.1) | 3.53** | <0.0001 (<0.0001) |
| | Exon 6 (899 bp) | 46 | 5 | 1.92 | 0.4 (0.4) | 1.27 | 0.8 (0.8) | 1.28* | 0.02 (0.02) |
| | Total (6041 bp) | 43 | 93 | 6.82** | 0.0001 (0.007) | 3.52* | 0.04 (0.07) | 3.39** | <0.0001 (<0.0001) |
| Capuchin monkeys | Exon 1 (1026 bp) | 31 | 8 | 3.86** | 0.001 (0.009) | 1.95 | 0.07 (0.08) | 2.94** | 0.0001 (<0.0001) |
| | Exon 3 (1072 bp) | 31 | 32 | 12.49** | <0.0001 (<0.0001) | 7.47** | <0.0001 (<0.0001) | 2.42** | 0.0009 (0.001) |
| | Exon 5 (1024 bp) | 31 | 27 | 11.06** | <0.0001 (<0.0001) | 6.60** | <0.0001 (<0.0001) | 2.41** | 0.0009 (0.001) |
| | Exon 6 (836 bp) | 31 | 21 | 11.30** | <0.0001 (<0.0001) | 6.29** | <0.0001 (0.0001) | 2.77** | 0.0001 (0.0001) |
| | Total (3958 bp) | 31 | 88 | 9.59** | <0.0001 (<0.0001) | 5.57** | <0.0001 (0.0002) | 2.74** | 0.0002 (0.0002) |

[a] Prob$\{\pi_{sim} > \pi_{obs}\}$, where $\pi_{sim}$ is the expected range of $\pi$ values in the simulation (fig. S2) using $\theta_A$ values and $\pi_{obs}$ is the observed $\pi$ values.

[b] Prob$\{\theta_{wsim} > \theta_{wobs}\}$, where $\theta_{wsim}$ is the expected range of $\theta_w$ values in the simulation (fig. S3) using $\theta_A$ values and $\theta_{wobs}$ is the observed $\theta_w$ values.

[c] Prob$\{D_{sim} > D_{obs}\}$, where $D_{sim}$ is the expected range of Tajima's *D* values in the simulation (fig. S4) using $\theta_A$ values and $D_{obs}$ is the observed Tajima's *D* values.

* Significant at 5% level.

** Significant at 1% level.

The *p* values in parentheses are those when $\rho$ was set to zero.

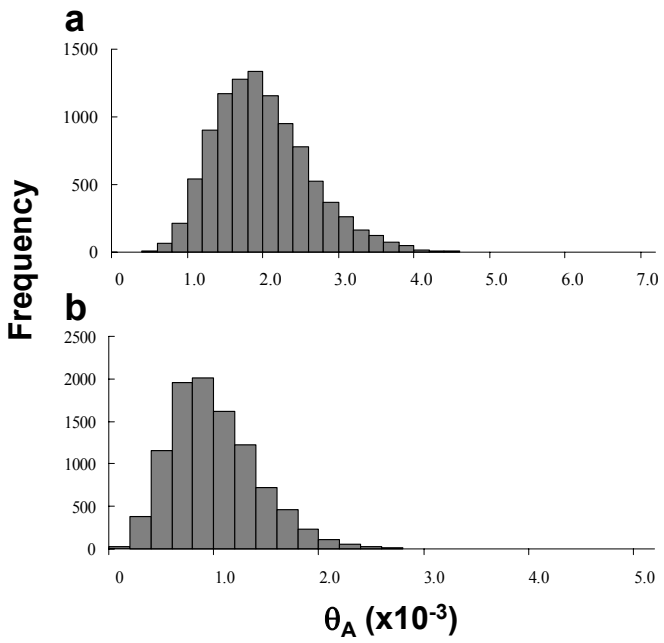


**Figure S1.** Samples of the posterior distribution of 10000 $\theta_A$ values in spider monkeys (a) and in capuchin monkeys (b) obtained by a coalescence simulation conditional on the observation of $\pi$ in the neutral references. Bin size is set to $0.2 \times 10^{-3}$. These distributions essentially represent the likelihood functions of $\theta_A$. The distributions are completed in the ranges given, $0 \sim 7.0 \times 10^{-3}$ for spider monkeys and $0 \sim 5.0 \times 10^{-3}$ for capuchin monkeys, ensuring the complete coverage of $\theta_A$ realizing $\pi$ to be within $\pm 1\%$ of the observed average of $\pi$ values among the five neutral references in the simulation.

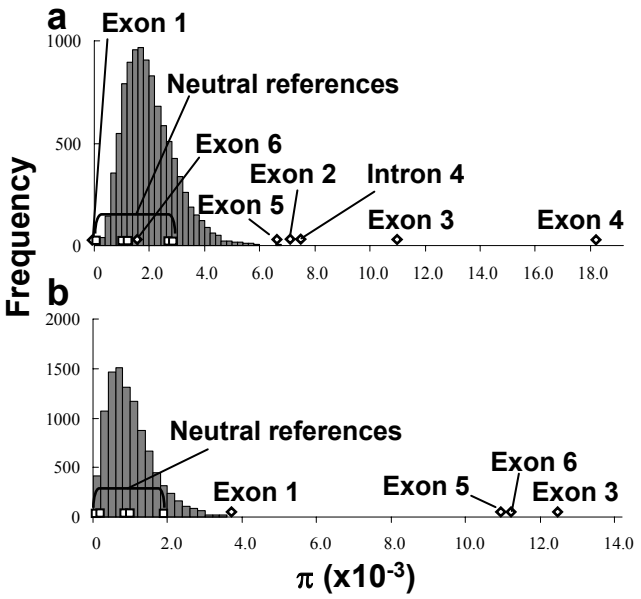


**Figure S2.** Expected distribution of 10000 π values in the sample size of L-M opsin gene of spider monkeys (a) and capuchin monkeys (b) obtained by a coalescence simulation based on the distribution of $\theta_A$ shown in fig. S1. Bin size is set to 0.2 x $10^{-3}$. The observed π values are plotted for the L-M opsin gene regions and the neutral references.

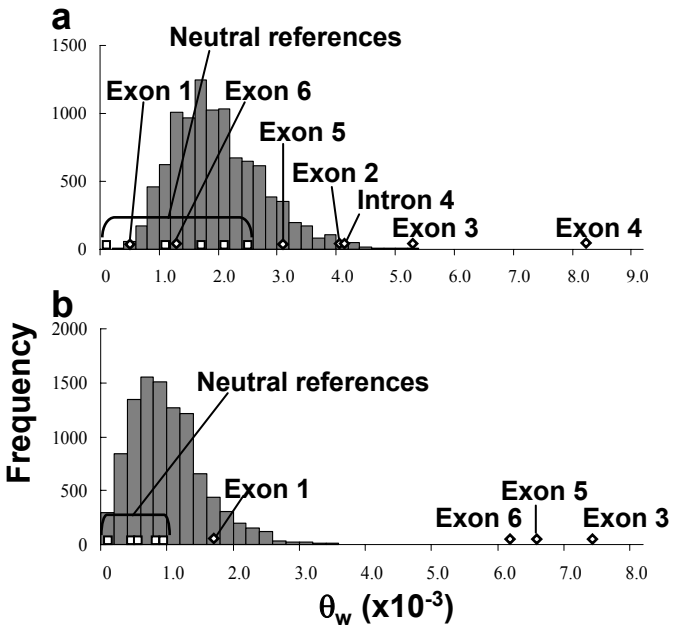


**Figure S3.** Expected distribution of 10000 $\theta_w$ values in the sample size of L-M opsin gene of spider monkeys (a) and capuchin monkeys (b) obtained by a coalescence simulation based on the distribution of $\theta_A$ shown in fig. S1. Bin size is set to 0.2 x $10^{-3}$. The observed $\theta_w$ values are plotted for the L-M opsin gene regions and the neutral references.

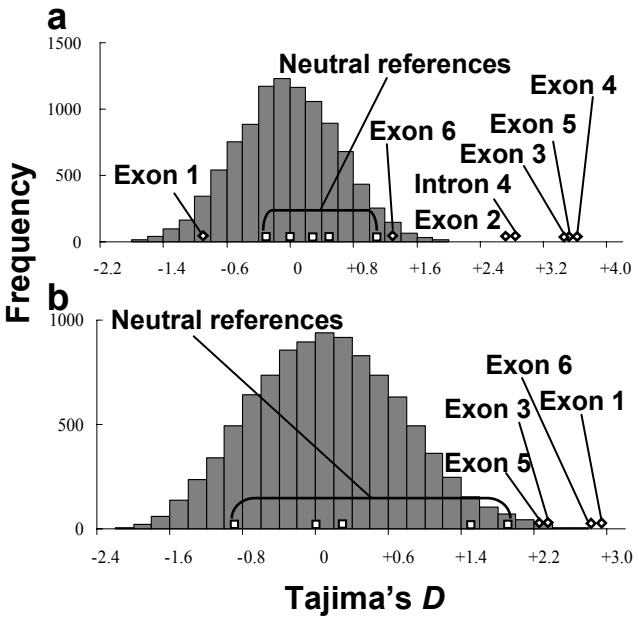


**Figure S4.** Expected distribution of 10000 Tajima's *D* values in the sample size of L-M opsin gene of spider monkeys (a) and capuchin monkeys (b) obtained by a coalescence simulation based on the distribution of $\theta_A$ shown in fig. 1. Bin size is set to $0.2 \times 10^{-3}$. The observed Tajima's *D* values are plotted for the L-M opsin gene regions and the neutral references.