Adaptive Evolution of Color Vision Genes in Higher Primates

Song-Kun Shyue, David Hewett-Emmett, Harry G. Sperling, David M. Hunt, James K. Bowmaker, John D. Mollon, Wen-Hsiung Li*

The intron 4 sequences of the three polymorphic alleles at the X-linked color photopigment locus in the squirrel monkey and the marmoset reveal that the alleles in each species are exceptionally divergent. The data further suggest either that each triallelic system has arisen independently in these two New World monkey lineages, or that in each species at least seven deletions and insertions (14 in the two species) in intron 4 have been transferred and homogenized among the alleles by gene conversion or recombination. In either case, the alleles in each species apparently have persisted more than 5 million years and probably have been maintained by overdominant selection.

Humans, apes, and Old World monkeys have trichromatic vision because they possess an autosomal gene that encodes a blue light–sensitive pigment and at least two X-linked genes that encode red- and green-sensitive pigments (*1*). In contrast, New World monkeys (NWMs) possess only one X-linked and one autosomal color photopigment gene. However, in squirrel monkeys (*Saimiri sciureus*) and marmosets (*Callithrix jacchus jacchus*), this single X-linked locus has three polymorphic alleles (*2–6*), which code for different photopigments whose sensitivities lie within the spectral range delimited by the human red and green pigments. Thus, although a male NWM is dichromatic, a heterozygous female is trichromatic, which might be the selective advantage that maintains these alleles (*3, 4, 7*). The origin of the triallelic systems in the two NWM lineages and their relation to the human red and green photopigment genes are not clear. To address these questions, we have sequenced portions of the various alleles (*8*).

The three squirrel monkey alleles encode proteins with spectral sensitivity maxima near 534, 550, and 561 nm (*4, 9*) and are denoted P534, P550, and P561, respectively. The spectral sensitivities of P534 and P561 are, respectively, close to those of human green (530 nm) and red (560 nm) photopigments, which are denoted P530 and P560. Sequence analysis of exon 5 was used to identify the three squirrel monkey alleles (*10*). The three marmoset alleles encode proteins with spectral sensitivity maxima near 543, 556, and 563 nm and are denoted P543, P556, and P563, respectively (*11*).

The pairwise divergences in intron 4 are ~2% between the squirrel monkey alleles and 3 to 4% between the marmoset alleles (Table 1). These are extremely high values for intraspecies sequence divergences, given that the average value for the divergence between two allelic sequences in humans is only 0.1% (*12*) and that the divergences between many different NWM species (or Old World monkey species) are less than 2% in the noncoding regions of the ε-globin gene (*13*). These high intraspecies divergences notwithstanding, the divergences between the squirrel monkey and marmoset alleles (5.8 to 6.5%) are significantly larger ($P < 0.05$). The divergences between the human and NWM sequences (about 12%) are in turn significantly larger ($P < 0.01$) than those between any pair of NWM alleles. The phylogenetic tree inferred by the neighbor-joining method (*14*) is the same as the species tree (Fig. 1A): Both the squirrel monkey alleles and the marmoset alleles diverged after the separation of the squirrel monkey and marmoset lineages. The same conclusion was reached by parsimony analysis. These analyses suggest that the marmoset and squirrel monkey alleles arose separately; note that the squirrel monkey alleles P534 and P550 differ in spectral peaks from the marmoset alleles P543 and P556.

We now consider insertions and deletions, which we refer to collectively as gaps (Fig. 1B). The presence of an *Alu* repeat in all squirrel monkey and marmoset alleles but not in the human red and green photopigment genes supports the hypothesis of separate origins for the NWM alleles and the two human genes. This distinction is further supported by the finding of 14 gaps [eight gaps of 1 base pair (bp) and the other six ranging in size from 2 to 24 bp] that distinguish the two human genes from the NWM alleles. The divergence of the squirrel monkey alleles from the marmoset alleles is supported by 14 unique gaps, seven shared by the three squirrel monkey alleles (four 1-bp deletions, one 2-bp deletion, one 15-bp deletion, and one 26-bp deletion) and seven shared by the three marmoset alleles (three 1-bp deletions, two 2-bp deletions, one 4-bp insertion, and one 35-bp deletion). It is more parsimonious to explain these gaps as resulting from separate origins for the multiallelic systems in the two NWMs than from a single origin (that is, the alleles arose in a common ancestor of the two species): Each gap in a species requires only one event (mutation) under the former hypothesis but requires one mutation in one of the three alleles and two transfers to the other two alleles under the latter; for the 14 unique gaps, the latter hypothesis requires 28 more events than does the former.

The relations among the three squirrel monkey alleles are uncertain because the branching order is supported by only 54% of the bootstrap replicates (Fig. 1A) and is in conflict with the findings that P534 and P561 share a 5-bp deletion and P561 and P550 share a 1-bp insertion (Fig. 1B). Therefore, it seems that the divergence among the three alleles was nearly a trichotomy. Among the marmoset alleles, P563 may have diverged before P556 and P543. This inference is supported by 70 or 92% of the bootstrap replicates by the

S.-K. Shyue, D. Hewett-Emmett, W.-H. Li, Human Genetics Center, School of Public Health, University of Texas, P.O. Box 20334, Houston, TX 77225, USA.
H. G. Sperling, Department of Ophthalmology and Visual Science, University of Texas Medical School, Houston, TX 77030, USA.
D. M. Hunt, Department of Molecular Genetics, Institute of Ophthalmology, University of London, London EC1V 9EL, UK.
J. K. Bowmaker, Department of Visual Science, Institute of Ophthalmology, University of London, London EC1V 9EL, UK.
J. D. Mollon, Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK.
*To whom correspondence should be addressed. E-mail: li@gsbs28.gs.uth.tmc.edu

Table 1. Number of nucleotide substitutions per 100 sites between intron 4 sequences [Sqm, squirrel monkey; Mar, marmoset; Hum, human; and X, human red and green genes, which are identical in intron 4 (*16*)]. Means and standard errors were estimated as in (*23*). Gaps are not included in the comparison.

Sequences	Sqm-P561	Sqm-P550	Sqm-P534	Mar-P563	Mar-P556	Mar-P543
Sqm-P550	2.01 ± 0.40					
Sqm-P534	1.93 ± 0.39	1.78 ± 0.37				
Mar-P563	6.50 ± 0.74	6.07 ± 0.71	6.40 ± 0.73			
Mar-P556	6.07 ± 0.71	5.97 ± 0.70	6.14 ± 0.71	3.93 ± 0.56		
Mar-P543	5.91 ± 0.70	5.90 ± 0.70	5.89 ± 0.70	3.84 ± 0.56	3.21 ± 0.51	
Hum-X	12.14 ± 1.05	12.21 ± 1.05	12.40 ± 1.06	12.75 ± 1.08	12.55 ± 1.06	12.58 ± 1.07

neighbor-joining and parsimony methods, respectively. However, it conflicts with the observation that P563 and P556 share a 23-bp deletion (Fig. 1B). Thus, the branching order for the three marmoset alleles is not certain.

The age of the NWM alleles was estimated with the assumption of a trichotomous divergence among the three alleles in each species. The average divergences in intron 4 are 1.9% between the three squirrel monkey alleles, 3.7% between the three marmoset alleles, and 6.1% between the two species. Assuming that the two NWM species diverged between 16.4 million and 19.0 million years ago (Ma) (*13*), we estimate that the three squirrel monkey alleles diverged between 5.1 and 5.9 Ma and that the three marmoset alleles diverged between 9.8 and 11.4 Ma. These values would be underestimates if gene conversion had occurred between any two alleles. These estimates, and those for apes and humans (*15*), suggest that ancient polymorphisms in color photopigments are common in primates.

Although it is more parsimonious to explain the intron 4 sequence data under the multiorigin hypothesis than under the single origin hypothesis, the latter cannot be ruled out because gene conversion (or recombination) might have occurred between sequences with an appreciable frequency. For example, gene conversion was proposed to explain the finding that the intron 4 sequences of the human red and green photopigment genes are identical (*16*), and deletion B5 or insertion L1 (Fig. 1B) should represent a transfer between two alleles. Under the single origin hypothesis, any one of the 14 gaps that existed in the common squirrel monkey–marmoset ancestor could not have been present on all three alleles, but gene conversion or recombination was needed to replace this gap by a sequence without this gap in one descendant lineage as the gap was spread to all three alleles in the other lineage. Of course, any gap that arose after separation of the two lineages would require only homogenization among the three alleles in one lineage.

To determine whether parallel evolution has occurred in amino acid sequences among the NWM and human X-linked photopigments, we examined amino acid positions 180, 277, and 285 (Fig. 2), which have been suggested to be critical for spectral tuning (9). Only under the assumption that the human and NWM systems have a single origin will no parallel change be required; that is, the Ser-Tyr-Thr and Ala-Phe-Ala alleles are assumed to have already existed before the human-NWM split. This scenario requires a Phe-to-Tyr change at position 277 in the common ancestor of the marmoset P556 and P543 alleles and an Ala-to-Thr change at position 285 in the squirrel monkey P550 allele. As noted above, the human and NWM systems probably have separate origins, which suggests that at least three parallel changes should have occurred: Ser to Ala at position 180, Tyr to Phe at position 277, and Thr to Ala at position 285. If the squirrel monkey and marmoset systems have separate origins (Fig. 2), then Ser-to-Ala changes at position 180 and Thr-to-Ala changes at position 285 of both species are required. Therefore, many parallel changes may have occurred at these critical positions, depending on the evolutionary relations among the human and NWM systems. Parallel changes (or convergent evolution) have previously been found at positions 180, 277, and 285 in the cave fish *Astyanax fasciatus* (*17*) and at position 277 in squids (*18*). In addition, the red photopigment (625 nm) of the American chameleon (*Anolis carolinensis*) also has Ser, Tyr, and Thr at the three critical sites (*19*).

Both the multiorigin and single origin scenarios for the NWM triallelic systems have intriguing evolutionary implications. If the multiorigin scenario is true, divergence between X-linked color photopigments has

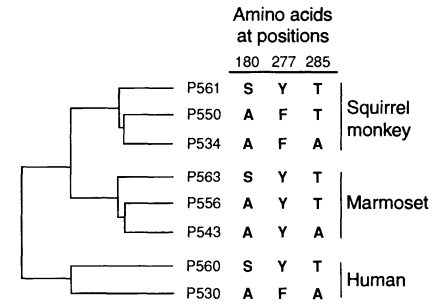
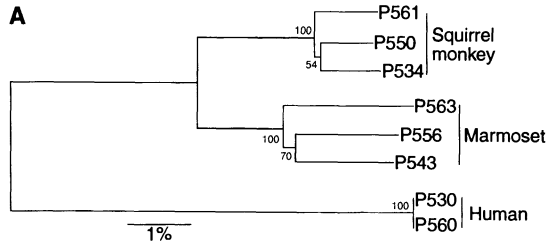


Fig. 2. Amino acids at three critical sites (180, 277, and 285) among eight primate color photopigments (A, Ala; F, Phe; S, Ser; T, Thr; and Y, Tyr).



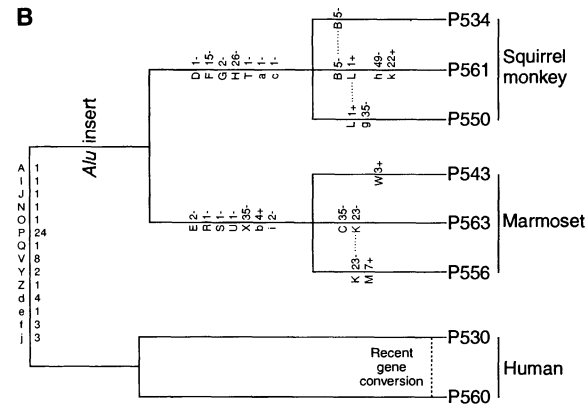


Fig. 1. (**A**) Phylogenetic tree derived from an analysis of intron 4 of the squirrel monkey alleles, the marmoset alleles, and the human red (P560) and green (P530) visual pigment genes. The tree was inferred by the neighbor-joining method; the number at each node denotes the proportion of 500 bootstrap replicates that supported the subset of sequences. The sequence divergences were computed as in (*23*). A parsimony analysis that excluded insertions and deletions and gave equal weights for all substitutions yielded the same tree topology and bootstrap proportions, with two exceptions: The bootstrap value for the P550-P534 clade was 41% instead of 54%, and that for the P556-P543 clade was 92% instead of 70%. (**B**) Gaps (insertions and deletions) supporting each branching node. The gaps are arranged alphabetically along the sequence alignment (available on request). Fourteen gaps plus one *Alu* repeat separate the two human genes from the squirrel monkey and marmoset alleles. These gaps are denoted by alphabetic location codes (see below); the size of each gap is given in base pairs. As there is no outgroup available for inferring the ancestral status, we cannot determine whether a gap was an insertion or a deletion and whether it occurred in the common ancestor of the NWM alleles or in the common ancestor of the two human genes. For the other gaps we can infer by the parsimony principle whether they were insertions (+) or deletions (−); for example, D1− means that the fourth gap (position 631 in the alignment; see below) was a deletion of 1 bp and occurred in the lineage leading to the squirrel monkey alleles. A dashed line denotes a gap shared by two alleles.

From sequence data on coding regions, we infer that the separation of the human red and green genes occurred before the divergence between the Old World monkey and human lineages but that a gene conversion in intron 4 between the two genes occurred in the recent past (*16*). The locations of the gaps in the 1942-bp alignment of intron 4 sequences are as follows: A, 32; *Alu*, 80 to 398; B, 401 to 405; C, 416 to 450; D, 631; E, 660 to 661; F, 671 to 685; G, 711 to 712; H, 780 to 805; I, 808; J, 820; K, 842 to 864; L, 878; M, 894 to 900; N, 914; O, 958; P, 970 to 993; Q, 1001; R, 1020; S, 1060; T, 1071; U, 1106; V, 1115 to 1122; W, 1256 to 1258; X, 1287 to 1321; Y, 1327 to 1328; Z, 1460; a, 1505; b, 1521 to 1524; c, 1531; d, 1605 to 1608; e, 1622; f, 1746 to 1748; g, 1766 to 1800; h, 1786 to 1834; i, 1792 to 1793; j, 1842 to 1844; and k, 1876 to 1897.

occurred between the red and green photopigment genes in the lineage leading to humans and at least twice between color photopigment alleles in NWMs. These repeated divergences and the required parallel amino acid substitutions would indicate adaptive evolution. In addition, the antiquity of the alleles implies that they have been maintained by balancing selection; otherwise, one or two of the alleles would have been lost through random drift (*20*). On the other hand, according to the single origin scenario, the triallelic system should be more ancient than the divergence of the two NWM species, which has been estimated at 16.4 to 19.0 Ma (*13*). Figure 1B suggests that at least seven gaps have been transferred and homogenized among the three alleles in each of the two species. This implies that the critical amino acids that define these alleles have been maintained in the presence of frequent homogenization events, which again suggests balancing selection. The type of selection is probably not minority advantage because the alleles in each species, with the possible exception of the marmoset P556, are maintained at high frequencies (*4*, *5*). One simple explanation is overdominant selection; trichromacy, which occurs in (female) heterozygotes, is thought to facilitate the detection of colored fruits against dappled foliage (*21*). The advantage of having three instead of two polymorphic alleles is an increase in the frequency of heterozygotes and thus the chance of overdominant selection. Another possible advantage of polymorphism is that monkeys with different spectral sensitivities may explore visually different environments (*3*). An additional advantage of the polymorphic system is that dichromats (males and homozygous females) detect color-camouflaged objects better than do trichromats (*22*). Thus, NWMs, which search for fruits cooperatively in groups, enjoy the advantages of both trichromacy and dichromacy (*6*).

REFERENCES AND NOTES

1. J. Nathans, D. Thomas, D. S. Hogness, *Science* **232**, 193 (1986); R. Ibbotson, D. M. Hunt, J. K. Bowmaker, J. D. Mollon, *Proc. R. Soc. London Ser. B Biol. Sci.* **247**, 145 (1992); M. Neitz and J. Neitz, *Science* **267**, 1013 (1995).
2. G. H. Jacobs, *Vision Res.* **24**, 1267 (1984).
3. J. D. Mollon, J. K. Bowmaker, G. H. Jacobs, *Proc. R. Soc. London Ser. B Biol. Sci.* **222**, 373 (1984).
4. G. H. Jacobs and J. Neitz, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2545 (1987).
5. D. S. Travis, J. K. Bowmaker, J. D. Mollon, *Vision Res.* **28**, 481 (1988).
6. M. J. Tovée, J. K. Bowmaker, J. D. Mollon, *ibid.* **32**, 867 (1992).
7. J. K. Bowmaker, G. H. Jacobs, J. D. Mollon, *Proc. R. Soc. London Ser. B Biol. Sci.* **231**, 383 (1987).
8. Intron 4 of each allele in each of the two species was amplified by the polymerase chain reaction (PCR); the primers and the experimental conditions used are available on request. Sequencing reactions were accomplished by dideoxynucleotide chain termination methods on double-stranded plasmid templates with T7 DNA polymerase (Sequenase kits, U.S. Biochemical). All sequences were determined in both directions by a combination of (i) direct sequencing with synthetic oligonucleotide primers and (ii) sequencing of exonuclease III–generated smaller subclones, which contain successively larger unidirectional deletions, with the Erase-a-Base system (Promega). At least three independent PCR clones were sequenced to avoid PCR errors. The GenBank accession numbers are X88888 through X88893.
9. G. H. Jacobs, J. Neitz, M. Neitz, *Vision Res.* **33**, 269 (1993).
10. M. Neitz, J. Neitz, and G. H. Jacobs [*Science* **252**, 971 (1991)] found that the three alleles in squirrel monkeys differ at two amino acid residues in exon 5 that are involved in spectral tuning: the amino acids for P534, P550, and P561 are, respectively, Phe, Phe, and Tyr at position 277 and Ala, Thr, and Thr at position 285. We had two male squirrel monkeys of Bolivian origin. One of them was determined by electroretinogram (ERG) to have a spectral peak at ~560 nm (that is, the P561 allele) and a preliminary ERG study suggested that the other had a spectral peak near 550 nm (that is, the P550 allele). PCR amplification of exon 5 of both alleles and subsequent nucleotide sequencing showed that the exon 5 sequences of the two alleles were identical to Neitz *et al.*'s sequences of P561 and P550, respectively. To identify P534, we obtained blood samples from 10 male squirrel monkeys of Bolivian origin from C. R. Abee of the University of South Alabama. After screening five DNA samples by PCR amplification and sequencing of exon 5 of the X-linked pigment gene, we found an individual with a sequence identical to Neitz *et al.*'s sequence of P534.
11. D. M. Hunt, A. J. Williams, J. K. Bowmaker, J. D. Mollon, *Vision Res.* **33**, 147 (1993).
12. W.-H. Li and L. A. Sadler, *Genetics* **129**, 513 (1991).
13. H. Schneider *et al.*, *Mol. Phylogenet. Evol.* **2**, 225 (1993).
14. N. Saitou and M. Nei, *Mol. Biol. Evol.* **4**, 406 (1987).
15. S. S. Deeb, A. L. Jorgensen, L. Battisti, L. Iwasaki, A. G. Motulsky, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7262 (1994); K. S. Dulai, J. K. Bowmaker, J. D. Mollon, D. M. Hunt, *Vision Res.* **34**, 2483 (1994).
16. S.-K. Shyue, L. Li, B. H.-J. Chang, W.-H. Li, *Mol. Biol. Evol.* **11**, 548 (1994).
17. R. Yokoyama and S. Yokoyama, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9315 (1990).
18. A. Morris, J. K. Bowmaker, D. M. Hunt, *Proc. R. Soc. London Ser. B Biol. Sci.* **254**, 233 (1993).
19. S. Kawamura and S. Yokoyama, *FEBS Lett.* **323**, 247 (1993).
20. M. Kimura and T. Ohta, *Genetics* **63**, 701 (1969); N. Takahata and M. Nei, *ibid.* **124**, 967 (1990).
21. J. D. Mollon, *J. Exp. Biol.* **146**, 21 (1989).
22. M. J. Morgan, A. Adam, J. D. Mollon, *Proc. R. Soc. London Ser. B Biol. Sci.* **248**, 291 (1992).

23. F. Tajima and M. Nei, *Mol. Biol. Evol.* **1**, 269 (1984).
24. We thank C. R. Abee and L. Williams for blood samples of squirrel monkeys. Supported by NIH grant EY10317.

20 April 1995; accepted 7 July 1995

Functional Characterization and Developmental Regulation of Mouse Telomerase RNA

María A. Blasco, Walter Funk, Bryant Villeponteau, Carol W. Greider*

Telomerase synthesizes telomeric DNA repeats onto chromosome ends de novo. The mouse telomerase RNA component was cloned and contained only 65 percent sequence identity with the human telomerase RNA. Alteration of the template region in vivo generated altered telomerase products. The shorter template regions of the mouse and other rodent telomerase RNAs could account for the shorter distribution of products (processivity) generated by the mouse enzyme relative to the human telomerase. Amounts of telomerase RNA increased in immortal cells derived from primary mouse fibroblasts. RNA was detected in all newborn mouse tissues tested but was decreased during postnatal development.

Telomerase is a ribonucleoprotein DNA polymerase that maintains telomere length by adding telomeric sequences onto chromosome ends (*1*). Human and mouse telomerases differ in both their functional properties and their regulation. Partially purified mouse telomerase adds predominantly only one repeat onto a telomeric primer in vitro, whereas the human enzyme adds hundreds of repeats under identical conditions and mixing extracts does not alter the processivity of either enzyme (*2*). In contrast to most normal human tissues, some normal mouse somatic tissues have detectable telomerase activity (*3*). Mouse cells can spontaneously immortalize in culture, whereas human cells rarely, if ever, spontaneously immortalize (*4*). The differential regulation of mouse and human telomerase may affect the ease of immortalization of mouse cells in culture (*3*).

To investigate the difference between the human and mouse enzymes, we cloned the mouse telomerase RNA component. A mouse genomic clone was identified (*5*) by hybridization to a 450-nucleotide (nt) probe from the transcribed region of the human telomerase RNA gene. The transcribed region of the mouse gene was 65% identical to the human telomerase RNA gene (*6*), which indicates that this clone might be the mouse telomerase RNA gene (Fig. 1) (*7*). The sequence identity in the transcribed region of the human and putative mouse RNA genes is significantly less than that

M. A. Blasco and C. W. Greider, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.
W. Funk and B. Villeponteau, Geron Corporation, Menlo Park, CA 94025, USA.

*To whom correspondence should be addressed.