The color vision of mammals is controlled by photosensitive proteins called opsins. Most mammals have dichromatic color vision, but hominoids and Old World (OW) monkeys enjoy trichromatic vision, having the blue-, green-, and red-sensitive opsin genes. Most New World (NW) monkeys are either dichromatic or trichromatic, depending on the sex and genotype. Trichromacy in higher primates is believed to have evolved to facilitate the detection of yellow and red fruits against dappled foliage, but the process of evolutionary change from dichromacy to trichromacy is not well understood. Using the parsimony and the newly developed Bayesian methods, we inferred the amino acid sequences of opsins of ancestral organisms of higher primates. The results suggest that the ancestors of OW and NW monkeys lacked the green gene and that the green gene later evolved from the red gene. The fact that the red/green opsin gene has survived the long nocturnal stage of mammalian evolution and that it is under strong purifying selection in organisms that live in dark environments suggests that this gene has another important function in addition to color vision, probably the control of circadian rhythms.

Introduction

The color vision of mammals is controlled by photopigments, each of which consists of a protein called opsin and a chromophore 11-cis retinal. The diversity of color vision in mammals is due to variation in the number and sequence of opsin genes. Most mammals have a short wavelength (blue)-sensitive opsin gene and a middle wavelength (green)- to long wavelength (red)-sensitive opsin gene, so that they have dichromatic color vision. In higher primates, however, trichromatic color vision evolved in two different ways. Hominoids and Old World (OW) monkeys have three different color vision genes: one autosomal gene encoding the blue-sensitive opsin and two X-linked genes encoding the green- and red-sensitive opsins (Nathans, Thomas, and Hogness 1986). Most New World (NW) monkeys have only one X-linked locus in addition to the autosomal blue gene locus, but the X-linked locus is polymorphic and contains three different alleles (Mollon, Bowmaker, and Jacobs 1984; Neitz, Neitz, and Jacobs 1991). Therefore, all males are dichromatic, whereas females are either dichromatic or trichromatic depending on the genotype. Jacobs et al. (1996) recently showed that one genus of NW monkeys (Alouatta; howler monkey) has two X-linked opsin genes, but this is probably due to a recent gene duplication confined to this genus and does not change the general picture of trichromacy due to polymorphism in NW monkeys. The blue and red/green genes seem to have diverged more than 500 MYA, but the green and red genes apparently diverged only after OW monkeys separated from NW monkeys (Nathans, Thomas, and Hogness 1986; Yokoyama and Yokoyama 1989). Trichromacy in higher primates is believed to have evolved to facilitate the detection of yellow and red fruits against dappled foliage (Mollon 1991). However, the process of evolutionary change from dichromacy to trichromacy is not well understood (Bowmaker 1991; Mollon 1991; Jacobs 1993). This is partly because the color vision of ancestral species of higher primates is unknown. Although a number of authors have speculated on the origin of red and green opsins (Yokoyama and Yokoyama 1990; Jacobs 1993; Wenderickx et al. 1993), no serious study has been done on this subject.

The purpose of this paper is to infer the color vision of ancestral higher primates and study the evolution of color vision. Since the color vision in mammals is entirely determined by the amino acid sequences of opsins, the problem becomes how to infer the opsin sequences of ancestral higher primates. Ancestral protein sequences can be statistically inferred from present-day sequences by various methods. Computer simulation (Zhang and Nei 1997) has shown that when the level of sequence divergence is low, the statistical inference is quite accurate. In general, once the ancestral sequence of a protein is inferred, its function is studied by producing the ancestral protein artificially (e.g., Jermann et al. 1995; Chandrasekharan et al. 1996). In the case of the red and green opsins of higher primates, however, the photosensitivity of an opsin is determined primarily by three critical amino acids, as will be mentioned later. Therefore, we can infer the color vision of ancestral primate organisms without producing ancestral proteins.

In the course of this study, we reached the conclusion that some opsin genes probably have another important function in addition to color vision. We will first discuss the color vision of ancestral higher primates and then provide evidence supporting the view that the green/red genes have another function.

Material and Methods

Background Information

The spectral sensitivity of opsins is measured by the maximum absorption wavelength ($\lambda_{\text{max}}$), which is 561 nm and 530 nm for the human red and green opsins, respectively. It is now well established that this difference in $\lambda_{\text{max}}$ between the red and green opsins is primarily controlled by three critical amino acids at sites
parts of exons 3, 4, and 5. However, six of the seven effects on $A_{\text{max}}$, it is possible to infer the color vision by examining the 126 sites.

Since amino acid site 116 has only a minor role in determining the difference in $A_{\text{max}}$ between the red and green genes, we decided to use the human sequences as representatives of hominoid sequences. Similarly, the pairwise Jukes-Cantor distance for hominoid species was in the range of 0.8%-2.4% for the red gene and 1.3%-3.5% for the green gene. Therefore, we decided to use the human sequences as representatives of hominoid sequences. Similarly, the pairwise Jukes-Cantor distance for hominoid species was in the range of 0.8%-2.4% for the red gene and 1.3%-3.5% for the green gene. Therefore, we decided to use the human sequences as representatives of hominoid sequences.

The ancestral amino acid sequences of the opsins were first inferred by Zhang and Nei’s (1997) distance-based Bayesian method, which is a slight modification of Yang, Kumar, and Nei’s (1995) likelihood-based Bayesian method. Computer simulation (Zhang and Nei 1997) has shown that this method gives more accurate inference than the commonly used parsimony method (e.g., Maddison and Maddison 1992) and is as good as the more time-consuming likelihood-based method. However, the parsimony method and the likelihood-based method are less accurate than the distance-based method. Therefore, we decided to use the human sequences as representatives of hominoid sequences. Similarly, the pairwise Jukes-Cantor distance for the three OW monkeys was in the range of 0.8%-1.3% for the red gene and 1.1%-2.7% for the green gene. Therefore, the macaque sequences were chosen as the representatives of the OW monkey sequences in the primary analysis. The phylogenetic relationships among the NW monkey opsins (three species each with three alleles) are unclear (e.g., Shyue et al. 1995), so we analyzed them species by species rather than simultaneously.

### Table 1

<table>
<thead>
<tr>
<th>Sequence Type</th>
<th>Red $A_{\text{max}}$ (nm)</th>
<th>Green $A_{\text{max}}$ (nm)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hominoids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human (Homo sapiens)</td>
<td>561</td>
<td>530</td>
<td>Neitz, Thomas, and Hogness (1986)</td>
</tr>
<tr>
<td>Pygmy chimpanzee (Pan paniscus)</td>
<td>560–565</td>
<td>530–535</td>
<td>Deeb et al. (1994)</td>
</tr>
<tr>
<td>Gorilla (Gorilla gorilla)</td>
<td>560–565</td>
<td>530–535</td>
<td>Dulai et al. (1994), Deeb et al. (1994)</td>
</tr>
<tr>
<td>Orangutan (Pongo pygmaeus)</td>
<td>560–565</td>
<td>530–535</td>
<td>Deeb et al. (1994)</td>
</tr>
<tr>
<td>Old World monkeys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macaque (Macaca fascicularis)</td>
<td>563</td>
<td>533</td>
<td>Dulai et al. (1994)</td>
</tr>
<tr>
<td>Diana monkey (Cercopithecus diana)</td>
<td>563</td>
<td>533</td>
<td>Dulai et al. (1994)</td>
</tr>
<tr>
<td>Talapoin monkey (Miopithecus talapoin)</td>
<td>563</td>
<td>533</td>
<td>Dulai et al. (1994)</td>
</tr>
<tr>
<td>New World monkeys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marmoset (Callithrix jaccus jaccus)</td>
<td>563, 556</td>
<td>543</td>
<td>Hunt et al. (1993)</td>
</tr>
<tr>
<td>Squirrel monkey (Saimiri sciureus)</td>
<td>561</td>
<td>547, 532</td>
<td>Neitz, Neitz, and Jacobs (1991)</td>
</tr>
<tr>
<td>Tamarin (Saguinus fuscicolis)</td>
<td>562, 556</td>
<td>541</td>
<td>Neitz, Neitz, and Jacobs (1991)</td>
</tr>
<tr>
<td>Outgroup species used</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goat (Capra hircus)</td>
<td>555</td>
<td></td>
<td>Unpublished data</td>
</tr>
<tr>
<td>Chicken (Gallus gallus)</td>
<td>571</td>
<td></td>
<td>Okano et al. (1992)</td>
</tr>
<tr>
<td>American chameleon (Anolis carolinensis)</td>
<td>561</td>
<td>521</td>
<td>Kawamura and Yokoyama (1993)</td>
</tr>
<tr>
<td>Gecko (Gecko gecko)</td>
<td>561</td>
<td>521</td>
<td>Kojima et al. (1992)</td>
</tr>
</tbody>
</table>

Note: Sequences less than 100 codons long are not included.

- These values are approximate. The $A_{\text{max}}$ value for the same organism varies with investigator to some extent, but the variation is usually within 5 nm.
- Color designation is somewhat arbitrary.
- Only amino acid sequences are available.
- This opsin is normally linked with 11-cis-3,4-dehydroretinal. To compare its $A_{\text{max}}$ with those of primate opsins, the $A_{\text{max}}$ is calculated under the assumption that it is reconstituted with 11-cis-retinal.

Sequences Used

All available sequences of primate red and green opsin genes are listed in table 1, together with the sequences of other higher vertebrates which were used as outgroup. The human red and green opsins have 364 amino acids, which are encoded by six exons interrupted by five introns. Unfortunately, many available sequences are not of full length. For example, the OW monkey sequences are only 126 amino acids long, comprising parts of exons 3, 4, and 5. However, six of the seven sites that determine the difference in $A_{\text{max}}$ between the human red and green opsins are located in this 126-codon region, the remaining one (site 116) being located in exon 2. Since amino acid site 116 has only a minor effect on $A_{\text{max}}$, it is possible to infer the color vision by examining the 126 sites.

The orthologous opsin gene sequences of hominoids are closely related to one another. The pairwise Jukes-Cantor distance for hominoid species was in the range of 0.8%-2.4% for the red gene and 1.3%-3.5% for the green gene. Therefore, we decided to use the human sequences as representatives of hominoid sequences. Similarly, the pairwise Jukes-Cantor distance for the three OW monkeys was in the range of 0.8%-1.3% for the red gene and 1.1%-2.7% for the green gene. Therefore, the macaque sequences were chosen as the representatives of the OW monkey sequences in the primary analysis. The phylogenetic relationships among the NW monkey opsins (three species each with three alleles) are unclear (e.g., Shyue et al. 1995), so we analyzed them species by species rather than simultaneously.

### Phylogenetic Analysis

Since the divergences among the primate opsin genes are fairly low, we reconstructed the phylogeny of opsin genes by using the neighbor-joining method (Saitou and Nei 1987) with Jukes-Cantor distances for nucleotide sequences. The amino acid sequences were not very informative in this case. The tree was tested by the bootstrap method (Felsenstein 1985) with 1,000 replications and the interior branch test (Zhou and Nei 1992). The computer software package MEGA (Kumar, Tamura, and Nei 1993) was used in these analyses.

### Ancestral Sequence Inference

The ancestral amino acid sequences of the opsins were first inferred by Zhang and Nei’s (1997) distance-based Bayesian method, which is a slight modification of Yang, Kumar, and Nei’s (1995) likelihood-based Bayesian method. Computer simulation (Zhang and Nei 1997) has shown that this method gives more accurate inference than the commonly used parsimony method (e.g., Maddison and Maddison 1992) and is as good as the more time-consuming likelihood-based method. However, the parsimony method and the likelihood-
based Bayesian method were also used to check the results from the distance-based Bayesian method. The amino acid substitution model (JTT-f model) we used for the Bayesian method was a modified version of the JTT-f model (based on Dayhoff, Schwartz, and Orcutt 1978) and the equal-input model were also used to see the effect of different substitution models. The JTT-f model and the Dayhoff-f model are based on empirical amino acid substitution data for many different proteins and therefore represent an average substitution pattern for many proteins, whereas the equal-input model assumes that the probability of change of amino acid $i$ to amino acid $j$ is proportional to the frequency of amino acid $j$ in the sequences.

Results and Discussion

Ancestral Opsin Sequences of Higher Primates

Inference of ancestral sequences requires a reliable phylogenetic tree (topology) of the present-day sequences. Figure 1a shows the phylogenetic tree of the red and green opsin genes for some representative higher primates and outgroup species. This tree was constructed by using the nucleotide sequences for the 126-codon region of exons 3, 4, and 5 given in Figure 2. The tree topology seems to be reliable except for the branch between nodes $b$ and $c$, which is not statistically well supported. Nevertheless, this tree topology is consistent with the current view of evolution of color vision genes in higher primates (Bowmaker 1991; Shyue et al. 1995), and the nonprimate portion of the tree is well supported by other biological information.

We inferred the amino acid sequences of the opsins at all interior nodes using this tree. As mentioned earlier, only 126 amino acids were inferred for each ancestral opsin, and the accuracy of inference of an amino acid was evaluated by the Bayesian posterior probability (Yang, Kumar, and Nei 1995; Zhang and Nei 1997). The results obtained are presented in Figure 2. In general, the accuracy of inference is quite high; only a few inferences of amino acids of ancestral primates have posterior probabilities that are lower than 90%. Figure 1a shows the inferred ancestral amino acids at the three critical sites (180, 277, and 285). Node $d$ has S (serine), Y (tyrosine), and T (threonine) at the three positions and node $e$ has A (alanine), F (phenylalanine), and A, as expected. Node $g$ also has AYT with a probability of nearly 100%. Therefore, our inference of amino acid at site 180 is ambiguous, but this ambiguity is understandable because human red opsin has either $S$ or $A$ at this position (see below). We also used the likelihood-based Bayesian method to infer the ancestral opsin sequences and obtained the same results. Essentially the same results were obtained for both the distance- and likelihood-based methods when three different models of amino acid substitution (the equal-input, Dayhoff-f, and JTT-f models) were used. The parsimony method with an unrooted tree (Maddison and Maddison 1992; Yang, Kumar, and Nei 1995) gave the same amino acids as those obtained by the Bayesian method at sites 277 and 285, but at site 180, parsimony gave two possibilities (S or A) for nodes $a$, $b$, $c$, $f$, $h$, and $i$ and $S$ for node $d$ and $A$ for nodes $e$ and $g$. Unlike the Bayesian methods, parsimony does not use information about branch lengths and therefore does not give the probability of the inferred amino acids, but the results obtained are consistent with those from the other two methods.

In higher primates, the $\lambda_{\text{max}}$ value varies from 530 to 547 nm for green opsins and from 556 to 565 nm for red opsins. We call the marmoset allele P556 a red gene, because human populations are polymorphic for alleles “SYT” and “AYT” at the red gene locus, and the individuals with these alleles both have normal color vision (Nathans, Thomas, and Hogness 1986; Winderickx et al. 1992). In primate red opsins, the amino acid changes S$\rightarrow$A, Y$\rightarrow$F, and T$\rightarrow$A at the three critical sites are known to reduce $\lambda_{\text{max}}$ by approximately 5, 9, and 15 nm, respectively, and these effects are more or less additive (Neitz, Neitz, and Jacobs 1991). If we use this rule, we can infer that the ancestral organism of hominoids and OW and NW monkeys had a red opsin with $\lambda_{\text{max}} = 561$ or 556 nm without any green opsin. This inference is further supported by the identity of amino acids between the sequence at node $b$ and the goat red gene sequence at the four sites that slightly affect the $\lambda_{\text{max}}$ values of primate red and green opsins (fig. 2). It is interesting to note that there are color-vision-deficient people who do not have any green genes, and they are called deuteranopes. Examining the DNA extracted from the preserved eye tissue of the famous chemist John Dalton, Hunt et al. (1995) showed that he was a deuteranope with amino acids A, Y, and T at the three critical sites of his red opsin. For deuteranopes, the color spectra that look yellow, orange, and red to normal persons are all of the same hue. Our primate ancestors apparently had a deuteranopia-like color vision. It would be interesting to reconstruct the predicted proteins by site-directed mutagenesis and examine their detailed function, as was done with artiodactyl ribonucleases (Jermann et al. 1995).

As mentioned earlier, there are four more sites (116, 230, 233, and 309) that have a minor effect (1–4 nm) on the difference in $\lambda_{\text{max}}$ between the human red and green opsins. In vertebrates other than humans, however, the amino acids at these sites do not necessarily correspond to those of the human red and green opsins (fig. 2). Therefore, the ancestral amino acids at these four sites do not seem to have much biological meaning. In our analysis mentioned above, amino acid site 116 in exon 2 was not included because of the lack of data for OW monkeys. However, we were able to
Fig. 1.—Phylogenetic trees used and inferred amino acids of ancestral organisms (tree nodes) of higher primates at the three critical amino acid sites of red/green opsins. The number after P in the gene symbol refers to $\lambda_{\text{max}}$. a. Ancestral amino acids inferred under the hypothesis of independent origins of trichromacy in OW and NW monkeys. The numbers below each interior branch are the bootstrap probability value followed by the confidence probability value (in percentage). A = alanine, S = serine, F = phenylalanine, Y = tyrosine, and T = threonine. b, The phylogenetic relationships of the primate red and green opsins under the single-origin hypothesis of trichromacy in higher primates. c. Ancestral amino acids inferred under the single-origin hypothesis.

Infer some of the ancestral amino acids inferred excluding OW monkeys. (One hundred ninety-eight amino acid sites were used in this analysis.) The results obtained show that all ancestral amino acids we inferred are Y at site 116 and are the same as the amino acid (Y) of the human green opsin rather than that (S) of the human red opsin (fig. 2). However, they are the same as that (Y) of the goat red opsin. These results suggest that the amino acid at this site is not important for determining $\lambda_{\text{max}}$.

Although only one species from each group of hominoids, OW monkeys and NW monkeys, was used
in the above study, further analyses using other species from these groups did not change our results appreciably. The only differences observed were the posterior probabilities of S and A at site 180. The tree in figure la implies that trichromatic color vision evolved independently in NW and OW monkeys. However, Hunt et al. (1993) and Dulai et al. (1994) suggested that trichromacy might have evolved first as a polymorphic form before the divergence of NW and OW monkeys and that the duplicate red and green genes in OW monkeys evolved later as a result of unequal crossing over. In this hypothesis, a high degree of similarity of red and green alleles (genes) within species is attributed to sequence homogenization due to gene conversion or recombination. (Gene conversion apparently occurs in the introns of opsin genes, but the exon regions do not appear to be subject to frequent gene conversion; Shyue et al. 1994; Zhou and Li 1996). Under this hypothesis, the relationship of the primate opsin sequences would be something like that given in figure 1b. However, the branch lengths from nodes b to x, b to z, and x to y in figure 1b are unknown. We therefore assumed as a first approximation that these branch lengths are all 0 but that the total length from node b to each present-day sequence remains the same as in figure 1a (see figure 1c). Although this assumption appears to be extreme, the tree of figure 1c is actually intermediate between the trees of figure 1a and 1b. At any rate, using this tree we inferred ancestral opsin sequences at all interior nodes. However, the results obtained were essentially the same as those for figure 1a, although at site 180 amino acid A now had a probability of 96% for nodes a and b. The parsimony method gave the same amino acids as those obtained by the Bayesian methods at sites 277 and 285, whereas at site 180 it gave two possibilities (A or S) for nodes h and i, S for node d, and A for nodes a, b, and e. Therefore, the two evolutionary hypotheses give different probabilities of A at this position, but this does not alter our conclusion about the color vision of primate ancestors. Actually, we can
even use the tree in figure 1b to infer the ancestral sequences. When we used this tree, many branch length estimates became 0, but the ancestral amino acids obtained at the three critical sites were the same as those in figure 1c. These results show that our inference of ancestral amino acids is not affected seriously by gene conversion or recombination.

Figure 1 suggests that the ancestral species of the goat, chicken, American chameleon, and gekko also had opsins with sequence AYT or SYT. Since the Ac>S change at site 180 has a relatively minor effect on color vision, it is tempting to conclude that these ancestral species also had red genes. However, since the color vision of birds and reptiles may be affected by some other factors such as the type of retinal chromophore and colored oil droplets that are lodged in photoreceptor cells, this conclusion may not be justified. Even in some nonprimate mammalian species, \( \lambda_{\text{max}} \) is not solely determined by the three critical amino acid positions. For example, the red/green homologous opsin in the rat \textit{Rattus norvegicus} has AYT at the three critical sites (unpublished data), although the \( \lambda_{\text{max}} \) is 510 nm (Jacobs 1993). For these reasons, it seems premature to predict color vision of nonprimate ancestral organisms by the three critical positions only, but our inference of the amino acids of important sites 277 and 285 seems quite reliable.

Comparing the red and green genes from the human and the Mexican characid fish \textit{Astyanax fasciatus}. Yokoyama and Yokoyama (1990) suggested that the red genes in humans and fish evolved from the green gene independently through identical amino acid substitutions at a few sites. Jacobs (1993) also speculated that the prototype of higher primate red/green opsins had a \( \lambda_{\text{max}} \) of 543 nm. Similarly, the hypothesis that the red gene evolved from the green gene was supported by Winderickx et al. (1993). Our results suggest that the green gene evolved from the red gene unlike these speculations.

### Opsins May Have Another Important Function in Addition to Color Vision

Until relatively recently it was believed that most mammals lost color vision due to the nocturnal living condition in the early stage of mammalian evolution and that only higher primates regained color vision (Walls 1942; Wistow 1993). Recent studies of opsin genes indicate that the blue and red/green genes of primates have coexisted in the genome for more than 500 Myr and, thus, the red/green gene did not become extinct during the long nocturnal stage of mammalian evolution.

The characid fish \textit{A. fasciatus} is widely distributed in rivers of Mexico and part of Texas. However, there are populations of this species that live in dark environments of Mexican caves. These populations are believed to have entered into the caves during the Pleistocene, possibly 0.5–1 MYA (Barr 1968; Nei 1975). The fish in these caves are carotenoids and have no eyes or reduced eyes. Since there is no need for color vision for these fish, one would expect that their opsin genes have become nonfunctional. Yokoyama et al. (1995) published the DNA sequences of several alleles of these genes from cave and river fish. We therefore computed the numbers of synonymous (\( d_s \)) and nonsynonymous (\( d_N \)) nucleotide substitutions per site to examine whether the opsin genes are still functional and whether they are under purifying selection or not. The \( d_s \) and \( d_N \) values were computed by Nei and Gojobori’s (1986) method for each pair of polymorphic alleles from the cave and the river populations. The average values of \( d_s \)’s and \( d_N \)’s are presented in table 2. In both river and cave populations, the \( d_s/d_N \) ratio is much higher than 1, suggesting that the cave fish genes have been subject to purifying selection. Note also that the \( d_s/d_N \) ratio is of the same order of magnitude as that (9.4) for the comparison of the goat and human red genes. These results strongly suggest that the red and green genes in cave fish are still functional. Interestingly, the (partial) amino acid sequence of the red/green opsin of the blind mole rat \textit{Spalax ehrenbergi}, which is believed to have lived a subterranean way of life for many millions of years (Quax-Junken et al. 1985), is also found to be highly conserved (Argamaso et al. 1995). These observations indicate that a dark environment, where no color vision is necessary, does not inactivate opsin genes. The fact that the red/green gene did not become extinct in dark environments suggests that this gene probably has another important function. One such possible function is the control of circadian rhythms of various cellular and physiological processes. Note that, although there is no need for an organism to know day and night in dark environments, circadian rhythms or some kind of biological clock is necessary for gene expression, cellular differentiation, embryonic development, etc. (Sassone-Corsi 1994). In the American chameleon, the red/green gene is expressed not only in the retina but also in the pineal gland that controls circadian rhythms (Kawamura and Yokoyama 1997); and in mammals there is a strong indication that the red/green gene is involved in the control of circadian rhythms (Argamaso et al. 1995). Tosini and Menaker (1996) also showed that retinas of the golden hamster exhibit circadian rhythms of melatonin synthesis. It is possible that the red/green gene is another example of “gene sharing” (Piattigorsky and Wistow 1989).
Acknowledgments

This work was supported by NIH grants (M.N. and S.Y.) and an NSF grant (M.N.).

LITERATURE CITED


MARCY K. UYENOYAMA, reviewing editor

Accepted February 14, 1997