

*Letter to the Editor***Frequent Gene Conversion Between Human Red and Green Opsin Genes****Zhongming Zhao, David Hewett-Emmett, Wen-Hsiung Li**

Human Genetics Center, School of Public Health, University of Texas, P.O. Box 20334, Houston, TX 77225, USA

Received: 29 September 1997 / Accepted: 29 September 1997

**Abstract.** To study the evolution of human X-linked red and green opsin genes, genomic sequences in large regions of the two genes were compared. The divergences in introns 3, 4, and 5 and the 3' flanking sequence of the two genes are significantly lower than those in exons 4 and 5. The homogenization mechanism of introns and the 3' flanking sequence of human red and green opsin genes is probably gene conversion, which also occurred in exons 1 and 6. At least one gene conversion event occurred in each of three regions (1, 3, and 5) in the sequences compared. In conclusion, gene conversion has occurred frequently between human red and green opsin genes, but exons 2, 3, 4, and 5 have been maintained distinct between the two genes by natural selection.

**Key words:** Red and green opsin genes — Gene conversion — Natural selection — Introns — Exons

A normal human possesses one red and at least one green opsin gene (Nathans et al. 1986; Vollrath et al. 1988; Drummond-Borg et al. 1989). These genes are tightly linked on the X chromosome with the red opsin gene at the 5' end of the cluster, and each has six exons and five introns. The gene duplication that gave rise to the red and green opsin genes is believed to have occurred before the divergence of Old World (OW) primates (about

25,000,000 years ago) because all OW monkeys, apes, and humans possess both genes. Given the antiquity of the gene duplication, the introns of the two genes should have diverged more than 8% (Li 1997). Yet, Shyue et al. (1994) found that introns 4 of the two genes from a male European were identical and introns 2 of the two genes diverged by only 0.3%. They proposed that introns 4 and 2 of these two genes were homogenized by gene conversion events. Later, Zhou and Li (1996) found that gene conversion events have also occurred in introns 4 of the two opsin genes in an Asian Indian, a chimpanzee, and a baboon. Gene conversion also apparently occurs often between exons of the red and green opsin genes in humans and OW monkeys (Ibbotson et al. 1992; Windrickx et al. 1993). Recently, large genomic regions of human red and green opsin genes have been sequenced by the Sanger Center, Cambridge, UK. From these data we can ask whether gene conversion events have occurred in many regions between the two genes. This investigation will lead to a better understanding of the extent of gene conversion between the two genes; for example, we should learn whether all noncoding regions of the two genes have been strongly homogenized. It will also lead to a better understanding of how residual differences between the two genes are located and have been maintained.

The DNA sequences of the red and green opsin genes were from human cosmids QC8B6 and cG1160 located in Xq28; the GenBank accession numbers are Z68193 and Z46936, respectively. The red opsin gene sequence is complete, but for the green opsin gene, only the region

**Table 1.** Mean and standard error of the number of nucleotide substitutions per 100 sites between human red and green pigment genes in exons, introns, and 3' flanking sequences<sup>a</sup>

	# of differences/ sequence length (bp)	Noncoding region (K)	Coding region	
			K <sub>S</sub>	K <sub>A</sub>
Intron 3	0/506	0.0 ± 0.0		
Exon 4	5/166		3.8 ± 3.0	3.3 ± 1.9
Intron 4	1/1554 <sup>b</sup>	0.1 ± 0.1		
Exon 5	10/240		4.7 ± 2.9	3.9 ± 1.5
Intron 5	2/2282	0.1 ± 0.1		
Exon 6	0/108		0.0 ± 0.0	0.0 ± 0.0
3' Flanking	2/1256	0.2 ± 0.1		
Exons 4, 5, and 6	15/514		3.4 ± 1.6	3.0 ± 1.0

<sup>a</sup> Note: These sequences are from the GenBank; the accession numbers are Z68193 and Z46936. The K value was computed by Kimura's two-parameter method (Kimura 1980), and the K<sub>S</sub> and K<sub>A</sub> values were computed by Li's method (Li 1993). Only about one-third of intron 3 was available for computation

<sup>b</sup> One single-nucleotide gap at 998 position

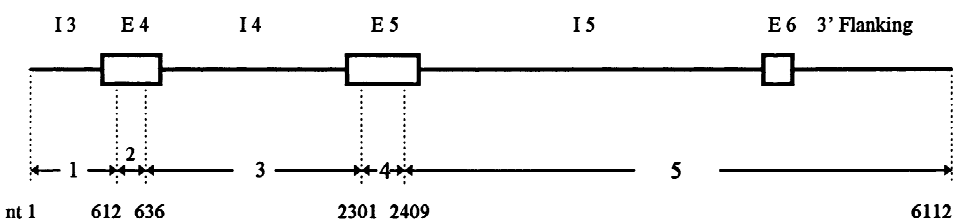
from the 3' part of intron 3 to the 3' flanking region is available. Therefore, we compared only the sequences of the 3' part of intron 3, exons 4, 5, and 6, introns 4 and 5, and the 3' flanking region of the red and green opsin genes. The total length is 6,112 bp. For intron 3 (506 bp) the two gene sequences are identical (Table 1). For intron 4 (1,554 bp), intron 5 (2,282 bp), and the 3' flanking sequence (1,256 bp), there are only one, two, and two nucleotide differences, respectively, between the two genes. In contrast, in exon 4 (166 bp) and exon 5 (240 bp), there are five and ten nucleotide differences between the red and green opsin genes, though the sequences of exon 6 (108 bp) of the two genes are identical. Shyue et al. (1994) found previously that the intron 4 sequences are identical in an European individual except for one single-nucleotide gap at position 32. We found one nucleotide difference and one single-nucleotide gap at position 998. The differences between the two studies may be due to different individuals or, possibly, PCR artifacts.

We then used Kimura's (1980) two-parameter method to compute the number of substitutions per site (K) between two intron sequences and Li's (1993) method to compute the number of substitutions per synonymous

site (K<sub>S</sub>) and per nonsynonymous site (K<sub>A</sub>) between two exon sequences (Table 1). Let us compare the divergences in introns 3, 4, and 5 and exons 4 and 5 because introns 3, 4, and 5 are adjacent to exons 3, 4, and 5, each of which contains amino acid residues critical to the spectral differences between the red and green opsin peptides (Winderickx et al. 1993). The percent divergences in introns 4 and 5 are both 0.1 ± 0.1 and that in intron 3 is 0.0 ± 0.0. All these values are significantly smaller than the synonymous substitutions per 100 sites (K<sub>S</sub>) and the nonsynonymous substitutions per 100 sites (K<sub>A</sub>) in exon 4 (3.8 ± 3.0 and 3.3 ± 1.9) and exon 5 (4.7 ± 2.9 and 3.9 ± 1.5). From these data, we can clearly see that introns 3, 4, and 5 of the red and green opsin genes have been completely or almost completely homogenized. The homogenization mechanism is probably gene conversion, not unequal crossing over, because the adjacent exons (4 and 5) have not been homogenized.

It is likely that gene conversion has also occurred in exons. In fact, exons 1 and 6 of the red and green opsin genes have been completely homogenized (Table 1; Nathans et al. 1986; Shyue et al. 1994). However, exons 2, 3, 4, and 5 each contain amino acid residues that are critical to the spectral differences between the red and green opsin peptides (Neitz et al. 1991; Merbs and Nathans 1992a,b, 1993; Asenjo et al. 1994). A gene conversion event in any of these exons may reduce the spectral sensitivity differences between the two opsins, may be selectively disadvantageous, and may be eliminated from the population (Shyue et al. 1994; Zhou and Li 1996). This is probably the reason that exons 2, 3, 4, and 5 of the two genes have been maintained distinct.

To infer the minimum number of gene conversion events that gave rise to the pattern of sequence similarity between the two opsin genes in these regions, we divide the red and green genes into five parts that have different degrees of divergence (Fig. 1). There are three regions (1, 3, and 5) that show a low degree of divergence. In the first region (positions 1 to 612), which includes intron 3 and the 5' end of exon 4, there is no difference between the two genes. In region 3 (637 to 2301), which includes the 3' end of exon 4, intron 4, and the 5' end of exon 5, there is only one nucleotide difference (position 851 in intron 4) in a total of 1,665 bp. In region 5 (2410 to 6112), which includes the 3' end of exon 5, intron 5,



**Fig. 1.** Partial opsin gene structure from intron 3 to the 3' flanking region. The sequence is divided into five parts (1 to 5) according to the divergences between the red and green opsin genes. The nucleotide positions are indicated at the bottom of the figure. I: intron, E: exon. The diagram is not drawn to scale.

exon 6, and the 3' flanking sequence, there are only four nucleotides different in a total of 3,703 bp. Clearly, at least one gene conversion event is required in each of these three regions because they are interrupted by regions 2 and 4, which have a much higher degree of divergence. In fact, the average divergence in regions 2 and 4 is 11.4 nucleotide differences per 100 sites, whereas the average divergence in regions 1, 3, and 5 is only 0.08 nucleotide differences per 100 sites, the former being 136 times higher than the latter.

In summary, we can conclude that gene conversion has occurred frequently between human red and green opsin genes, so that many regions in the two genes, e.g., introns 3, 4, and 5, and exon 6, have been almost completely homogenized. In contrast, exons 2, 3, 4, and 5 have been maintained distinct between the two genes by natural selection because they contain amino acid residues critical to the differences in spectral sensitivity between the two opsins.

*Acknowledgments.* This study was supported by an NIH grant.

## References

- Asenjo AB, Rim J, Oprian DD (1994) Molecular determinants of human red/green color discrimination. *Neuron* 12:1131–1138
- Drummond-Borg M, Deeb SS, Motulsky AG (1989) Molecular patterns of X chromosome-linked color vision genes among 134 men of European ancestry. *Proc Natl Acad Sci USA* 86:983–987
- Ibbotson R, Hunt DM, Bowmaker JK, Mollon JD (1993) Sequence divergence and copy number of the middle- and long-wave photopigment genes in Old World monkeys. *Proc R Soc Br* 247:145–154
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120
- Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* 36:96–99
- Li WH (1997) *Molecular evolution*. Sinauer, Sunderland, MA, pp 315–318
- Merbs SL, Nathans J (1992a) Absorption spectra of human cone pigments. *Nature* 356:433–435
- Merbs SL, Nathans J (1992b) Absorption spectra of the hybrid pigments responsible for anomalous color vision. *Science* 258:464–466
- Merbs SL, Nathans J (1993) Role of hydroxyl-bearing amino acids in differentially tuning the absorption spectra of the human red and green cone pigments. *Photochem Photobiol* 58:706–710
- 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
- Neitz M, Neitz J, Jacobs GH (1991) Spectral tuning of pigments underlying red-green color vision. *Science* 252:971–974
- Shyue SK, Li L, Chang BHJ, Li WH (1994) Intronic gene conversion in the evolution of human X-linked color vision genes. *Mol Biol Evol* 11(3):548–551
- Vollrath D, Nathans J, Davis RW (1988) Tandem array of human visual pigment genes at Xq28. *Science* 240:1669–1671
- Winderickx J, Battisti L, Hibiya Y, Motulsky AG, Deeb SS (1993) Haplotype diversity in the human red and green opsin genes: evidence for frequent sequence exchange in exon 3. *Hum Mol Genet* 2:1413–1421
- Zhou YH, Li WH (1996) Gene conversion and natural selection in the evolution of X-linked color vision genes in higher primates. *Mol Biol Evol* 13(6):780–783