Intronic Gene Conversion in the Evolution of Human X-Linked Color Vision Genes

Song-Kun Shyue, Li Li, Benny H.-J. Chang, and Wen-Hsiung Li
University of Texas Health Science Center at Houston

Human red and green visual pigment genes are X-linked duplicate genes. To study their evolutionary history, introns 2 and 4 (1,987 and 1,552 bp, respectively) of human red and green pigment genes were sequenced. Surprisingly, we found that intron 4 sequences of these two genes are identical and that the intron 2 sequences differ by only 0.3%. The low divergences are unexpected because the duplication event producing the two genes is believed to have occurred before the separation of the human and Old World monkey (OWM) lineages. Indeed, the divergences in the two introns are significantly lower than both the synonymous divergence (3.2% ± 1.1%) and the nonsynonymous divergence (2.0% ± 0.5%) in the coding sequences (exons 1-6). A comparison of partial sequences of exons 4 and 5 of human and OWM red and green pigment genes supports the hypothesis that the gene duplication occurred before the human-OWM split. In conclusion, the high similarities in the two intron sequences might be due to very recent gene conversion, probably during evolution of the human lineage.

In humans the three genes coding for the red, green, and blue color pigments have been cloned, and their coding sequences have been published (Nathans et al. 1986). The blue pigment gene is on chromosome 7, whereas the red and green pigment genes are tightly linked on the X chromosome, with the red pigment gene at the 5’ end of the tandem array (Vollrath et al. 1988). Normal individuals have one to five green pigment genes but only one red pigment gene (Drummond-Borg et al. 1989). Both the red and green pigment genes have six exons and five introns, and their coding sequences are very similar, i.e., only 25 differences of 1,092 nucleotides in the coding region (Nathans et al. 1986). Because humans, apes, and Old World monkeys (OWM) have one red pigment gene and at least one green pigment gene on the X chromosome, whereas New World monkeys (NWM) have only one X-linked pigment gene (Mollon et al. 1984; Jacobs and Neitz 1985, 1987), the duplication event that produced the red and green pigment genes probably occurred after the NWM-OWM split but before the human-OWM split.

To study the evolutionary history of human red and green pigment genes, we sequenced introns 2 and 4 of both genes. Unexpectedly, between the two genes the two intron sequences are significantly less divergent than the exon sequences. This is contrary to the usual situation in which introns are more divergent than exons and leads to the question of how this unusual situation arose.

Material and Methods

All the clones used in this study were isolated from human DNA libraries (Nathans et al. 1986) and were provided by Dr. J. Nathans, The Johns Hopkins University. For the human red pigment gene, the intron 2 sequences were obtained from lambda clones gJHN33 and gJHN53, and the intron 4 sequences were from lambda clone gJHN33. For the human green pigment gene, the intron 2 sequences were obtained from lambda clone gJHN43, and the intron 4 sequences were from both lambda clones gJHN43 and gJHN44.

Sequencing reactions were accomplished by the di-deoxynucleotide chain-termination methods (Sanger et al. 1977) on double-stranded templates, with T7 DNA polymerase (Sequenase kits, USB, Cleveland). Sequences were determined in both directions by a combination of (a) direct sequencing using synthetic oligonucleotide primers and (b) sequencing of exonuclease III (Exo III)-generated smaller subclones that contain successively larger unidirectional deletions, by using the Erase-a-Base kits (Promega).

To avoid potential confusion from the high similarity between intron sequences of the red and green pigment genes, all of the clones used for direct sequencing or Exo III deletion contained at least part of a flanking exon for verification. For both red and green pigment genes, exon 2 was included in the clone used for sequencing the 5’ end of intron 2; exon 3 was included in the clone used for sequencing the 3’ end of intron 2;
exons 4 and 5 were included in the clones used for sequencing intron 4.

For intron sequences the number of substitutions per site (K) between two sequences was computed by Kimura's (1980) two-parameter method. For coding sequences, the methods of Li (1993) and Pamilo and Bianchi (1993) were used to compute the number of substitutions per synonymous site (Ks) and per nonsynonymous site (Ka) between two sequences.

Results and Discussion

Nucleotide Sequences

With regard to intron 2, both intron sequences of the red and green pigment genes are 1,987 bp long and have one Alu element near the 3' end. The two intron sequences differ by only six nucleotides. With regard to intron 4, the red pigment gene has 1,552 bp. In the green pigment gene, clone gJHN43 has 1,552 bp, while clone gJHN44 has 1,551 bp because of a gap at position 32; the two sequences are otherwise identical. Because the genes are X-linked and because clones gJHN43 and gJHN44 were from the DNA library of a normal male individual, a simple explanation for the gap between the two clones is that they came from different green genes of the same person; since Nathans et al. (1986) found no difference in the exons, they assumed that they were overlapping clones derived from the same gene. It is interesting that the intron 4 sequence from clone gJHN43 is identical with the intron 4 sequence of the human red pigment gene, and it will be used in the following analysis. The four nucleotide sequences have been deposited in the EMBL data bank; accession numbers are X76096, X76097, X76098, and X76099.

Introns Are Less Divergent Than Exons

Table 1 shows the degree of divergence in introns and exons between human red and green pigment genes. The divergence in intron 2 is only 0.3%, which is smaller than both the synonymous (2.0%) and nonsynonymous divergence (1.5%) in exon 2 and the synonymous (5.6%) and nonsynonymous divergence (1.5%) in exon 3. The same is true for the comparison of intron 4 versus exons 4 and 5. Thus, both introns 2 and 4 are less divergent than arc their flanking exons. When all exons are considered together, the divergence between human red and green pigment genes is 3.2% ± 1.1% at synonymous sites and 2.0% ± 0.5% at nonsynonymous sites. Therefore, introns 2 and 4 (with only 0.3% and 0% divergences, respectively) are significantly less divergent than the coding sequences. Note that the 3.2% synonymous divergence is exceptionally low because the average synonymous divergence between human and OWM genes is >10% (Seino et al. 1992). Thus, it is remarkable that the divergence in introns 2 and 4 between human red and green pigment genes is even significantly <3.2%.

The above observation is contrary to the usual situation where coding sequences are better conserved than intron sequences (Li and Graur 1991). There are three possible interpretations for this observation: (i) the two introns contain functional sequences, such as coding sequences or control signals, in addition to splicing signals; (ii) the duplication event that produced the red and green pigment genes was so recent that not many mutations have yet accumulated in the introns; or (iii) gene conversion has occurred in both introns 2 and 4 in the recent past.

In order to know whether introns 2 and 4 contain functional sequences in addition to splicing signals, these introns are compared with introns 2 and 4 of the X-linked pigment gene in the squirrel monkey (an NWM, authors' unpublished sequences). The comparison reveals 11.8% divergence in intron 2 and 12.2% divergence in intron 4. These distances are within the range of the distances between humans and NWM obtained from intron sequences in other genes (Fitch et al. 1988; Spritz and Giebel 1988; Seino et al. 1992). In addition, there are insertions and deletions in both introns. Therefore, introns 2 and 4 of the primates red and green pigment genes are no more conservative than introns in other genes, and the high similarity between human red and green pigment genes in introns 2 and 4 is apparently not due to functional constraints.

To estimate the duplication date for human red and green pigment genes, exons 4 and 5 of the two genes are compared with the consensus partial sequences of exons 4 and 5 of the red and green pigment genes in six OWM sequences (Ibbotson et al. 1992) (table 2). For both exons 4 and 5, the Ka values between human and OWM red pigment genes or between human and OWM green pigment genes are considerably smaller than the values for all other comparisons. This observation implies that the human and OWM red pigment genes are more closely related to each other than either of them is to other genes and that the same is true for the human and OWM green pigment genes. In other words, the red and green pigment genes have diverged before the human-OWM split, as is commonly believed. Of course, we cannot rule out the possibility that the low Ka values in exons 4 and 5 between the two taxa are due to convergent evolution. However, the fact that all the Ks values

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Mean and Standard Error of the Number of Nucleotide Substitutions per 100 Sites between Human Red and Green Pigment Genes in Exons and Introns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (bp)</td>
<td>$K_s$</td>
</tr>
<tr>
<td>Exon 1</td>
<td>112</td>
</tr>
<tr>
<td>Exon 2</td>
<td>297</td>
</tr>
<tr>
<td>Intron 2</td>
<td>1,987</td>
</tr>
<tr>
<td>Exon 3</td>
<td>169</td>
</tr>
<tr>
<td>Exon 4</td>
<td>166</td>
</tr>
<tr>
<td>Intron 4</td>
<td>1,552</td>
</tr>
<tr>
<td>Exon 5</td>
<td>240</td>
</tr>
<tr>
<td>Exon 6</td>
<td>111</td>
</tr>
</tbody>
</table>

SOURCE.—Exon sequences are from Nathans et al. (1986).
to gene conversion (see below) or to exceptionally large peptides in their wave spectrum peaks (Neitz et al. 1991; genes diverged prior to the human-OWM split. The ex- might explain the higher \( K_s \) values in exons 4 and 5, statistical fluctuations owing to the short sequences in- Merbs and Nathans 1992; Williams et al. 1992; Win- inception is that in exon 4 the \( K_s \) value between human because amino acid differences in these exons are im- (with one exception, see below) are not significantly dif- ferent from one another confirms the view that the two genes diverged prior to the human-OWM split. The ex-ception is that in exon 4 the \( K_s \) value between human red and green pigment gene is low. This could be due to gene conversion (see below) or to exceptionally large statistical fluctuations owing to the short sequences involved.

**Gene Conversion**

A simple explanation for the higher similarities in introns than in exons is recent gene conversion in each of the two introns between the red and green pigment genes. Since the two introns are separated by intron 3 and by exons 3 and 4, which are more divergent than all other exons with the exception of exon 5, at least two conversion events are required to explain the high similarity in the two introns between human red and green pigment genes. As the two genes are tightly linked and have diverged only after the divergence between the hu- man and NWM lineages, i.e., probably <35 Mya, their sequences are still highly similar, so that chances for gene conversion to occur are high. We note that the degree of synonymous divergence between human red and green pigment genes is very low in exons 1, 2, 4, and 6 (table 1); in particular, no differences, either syn- onymous or nonsynonymous, have occurred in exons 1 and 6. Thus, gene conversion events probably have occurred in some of these exons. Elsewhere Ibbotson et al. (1992) and Balding et al. (1992) have reported evidence for gene conversion events in exons 4 and 5 of the red and green pigment genes in OWM.

Gene conversion can occur in exons as well as in introns. In the present case, however, gene conversion events in exons may be disadvantageous, and the resultant changes may be eliminated from the population, because such changes reduce the differences between the red and green pigment genes and thus also reduce the ability to distinguish between red and green colors. This might explain the higher \( K_s \) values in exons 4 and 5, because amino acid differences in these exons are important for the differences between the two pigment proteins in their wave spectrum peaks (Neitz et al. 1993; Merbs and Nathans 1992; Williams et al. 1992; Wind- derickx et al. 1992).

The high frequency of red-green or green-red fusion genes in human populations, ~16% of Caucasian males and 21% of African-American males (Drummond-Borg et al. 1989; Jorgensen et al. 1990), also suggests that during meiosis the red and green pigment genes may mispair frequently because of their high sequence similarity. Mispairing during meiosis increases the probability of gene conversion. On the other hand, high levels of similarity in introns facilitate mispairing and recom- bination, leading to production of hybrid genes.

**Acknowledgments**

We thank Dr. Jeremy Nathans for the human visual pigment gene lambda clones used in this study and Dr. David Hewett-Emmett for help and suggestions. This study was supported by NIH grants.

**LITERATURE CITED**


PAUL M. SHARP, reviewing editor

Received August 9, 1993

Accepted December 1, 1993