

The Evolution of Trichromatic Color Vision by Opsin Gene Duplication in New World and Old World Primates

Kanwaljit S. Dulai,¹ Miranda von Dornum,¹ John D. Mollon,² and David M. Hunt^{1,3}

¹Department of Molecular Genetics, Institute of Ophthalmology, University College London, London EC1V 9EL, UK;

²Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK

Trichromacy in all Old World primates is dependent on separate X-linked MW and LW opsin genes that are organized into a head-to-tail tandem array flanked on the upstream side by a locus control region (LCR). The 5' regions of these two genes show homology for only the first 236 bp, although within this region, the differences are conserved in humans, chimpanzees, and two species of cercopithecoid monkeys. In contrast, most New World primates have only a single polymorphic X-linked opsin gene; all males are dichromats and trichromacy is achieved only in those females that possess a different form of this gene on each X chromosome. By sequencing the upstream region of this gene in a New World monkey, the marmoset, we have been able to demonstrate the presence of an LCR in an equivalent position to that in Old World primates. Moreover, the marmoset sequence shows extensive homology from the coding region to the LCR with the upstream sequence of the human LW gene, a distance of >3 kb, whereas homology with the human MW gene is again limited to the first 236 bp, indicating that the divergent MW sequence identifies the site of insertion of the duplicated gene. This is further supported by the presence of an incomplete Alu element on the upstream side of this insertion point in the MW gene of both humans and a cercopithecoid monkey, with additional Alu elements present further upstream. Therefore, these Alu elements may have been involved in the initial gene duplication and may also be responsible for the high frequency of gene loss and gene duplication within the opsin gene array. Full trichromacy is present in one species of New World monkey, the howler monkey, in which separate MW and LW genes are again present. In contrast to the separate genes in humans, however, the upstream sequences of the two howler genes show homology with the marmoset for at least 600 bp, which is well beyond the point of divergence of the human MW and LW genes, and each sequence is associated with a different LCR, indicating that the duplication in the howler monkey involved the entire upstream region.

[The sequence data described in this paper have been submitted to GenBank under accession nos. AF155218, AF156715, and AF156716.]

Among the mammals, primates are unique in possessing trichromatic color vision. In humans and Old World primates, this is based on three classes of cone photoreceptor within the retina, each containing a different visual pigment. Visual pigments are composed of a chromophore, retinal, covalently bound by a protonated Schiff base to a protein moiety, opsin. Therefore, spectral differences between primate pigments are entirely attributable to differences in the amino acid sequence of the opsin proteins.

The spectral peaks (λ_{\max}) of the three visual pigments in primates are in the violet [short wave (SW)], green [middle wave (MW)], and yellow-green [long wave (LW)] regions of the spectrum. In Old World primates, the SW opsin is encoded by an autosomal gene and the MW and LW opsins by separate genes on the X chromosome (Nathans et al. 1986; Ibbotson et al.

1992). Sequence comparison between these two genes in different Old World species (Ibbotson et al. 1992) indicates that they arose by gene duplication at the base of the Old World lineage ~40 mya (Goodman et al. 1998; Hunt et al. 1998). The two genes show substantial sequence identity throughout their coding (Nathans et al. 1986) and intronic (Shyue et al. 1995; Zhao et al. 1998) regions. They are organized into a head-to-tail tandem array (Nathans et al. 1986; Vollrath et al. 1988; Feil et al. 1990), which is bound on the upstream side by a so-called locus control region (LCR) (Wang et al. 1992), the presence of which is critical for the expression of either gene (Nathans et al. 1989) and may be involved in the selective expression of one or other gene in particular cone photoreceptors (Shaaban and Deeb 1998).

The situation in New World monkeys is much more variable. The nocturnal monkey *Aotus trivirgatus* lacks a functional SW gene and is monochromatic (Ja-

³Corresponding author.
E-MAIL d.hunt@ucl.ac.uk; FAX 0171 608 6863.

cobs et al. 1996a). Most New World species exhibit a trichromacy that is based on only two opsin genes, an autosomal SW gene as in Old World primates, and a polymorphic X-linked MW/LW gene with up to three allelic forms that encode pigments with differing λ_{\max} values. Male monkeys combine the SW gene with just one of the different allelic forms of the X-linked gene and can generate only two cone pigments. Therefore, they are dichromats. In contrast, those females that inherit a different form of the MW/LW gene from each parent possess trichromatic vision, as X-inactivation will ensure that only one allele is expressed per cell (Mollon et al. 1984). Therefore, there is no need for an additional mechanism for the selective expression of the different allelic forms of the polymorphic X-linked gene in different cone photoreceptors.

An exception to this polymorphism-based trichromacy of New World primates has been identified by Jacobs et al. (1996b) in the howler monkey *Alouatta seniculus* and *Alouatta caraya*, where separate MW and LW genes are present, thus conferring full trichromacy in both males and females. Sequence analysis of the two genes indicates that this is a more recent duplication than in Old World primates (Hunt et al. 1998), which occurred after the separation of the Old World and New World lineages. For a functional trichromacy to be present in this genus, a system for the selective expression of either the MW or the LW gene in particular photoreceptors, is now required.

In this paper we have addressed the question of the origin of the duplication in Old World primates and in the howler monkey. The original report of the sequence of the human MW and LW genes by Nathans et al. (1986) included ~450 bp of the upstream sequence of both genes. This region shows only 14 single nucleotide substitutions in the first 236 bp, but then diverges completely, indicating that this is the point of insertion of the duplicated opsin gene. To confirm this, we have sequenced the upstream region of the single X-linked gene in the marmoset *Callithrix jacchus*, a New World monkey with a polymorphism-based trichromacy. The absence of any opsin gene duplication in this species means that the upstream sequence will reflect the ancestral arrangement. We have also examined the conservation of substitution within the region of homology of the MW and LW genes in humans, chimpanzee, and in two species of Old World cercopithecoid monkeys.

The presence of two distinct upstream regions to the X-linked opsin genes in the howler monkey has already been established (Kainz et al. 1998). We have extended this sequence across the region of divergence in the Old World primate genes and also into the LCR. From these sequences, we have been able to establish that the opsin gene duplication in this genus is very different from that in Old World primates.

RESULTS

The 5'-Flanking Regions of the Separate MW and LW Opsin Genes in Old World Primates

The first gene in the opsin gene array on the X chromosome is the LW gene and there is a distance of ~3.5 kb between the upstream LCR and the transcription start site for this gene [accession nos. S44029 (Wang et al. 1992) and Z68193 (Sanger Centre data bank 1995)]. Within this 5'-flanking sequence, a number of repeat elements are present including an *Alu* at -1991 to -2232, three L1 elements at -1360 to -1622, -1781 to -1970, and -2261 to -2341, and a MER12 element at -2438 to -2617. In the case of the MW gene, Nathans et al. (1986) originally published 450 bp of the human 5'-flanking region; this was extended by ~700 bp to -1185 by Dulai (1996) and further extended to -2513 by Hanna et al. (1997; accession no. U93721). A dot matrix plot of the first 1000 bp of the 5'-flanking regions of the MW and LW genes (Fig. 1) confirms that homology is restricted to the first 236 bp. Further analysis of the 5'-flanking region of the human MW gene shows that it contains four *Alu* repeat elements (Table 1), a partial repeat of 120 bp at position -234 immediately adjacent to the point of sequence divergence between the MW and LW 5' flanking regions, and three more complete elements further upstream. To establish whether these *Alu* elements are also present in the upstream region of the MW gene of another Old World primate, a fragment of ~750 bp of the upstream region of the MW gene in the diana monkey

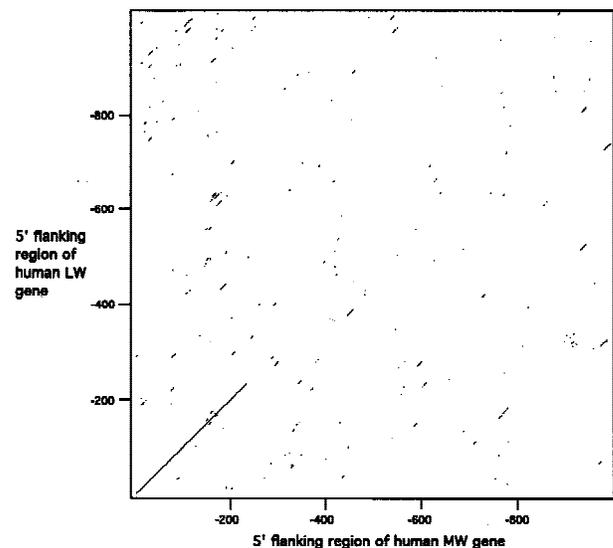


Figure 1 Dot matrix plot of human MW and LW upstream region. The parameters for comparison were set at 60% identity for a sliding window of 10 nucleotides. The presence of a diagonal line that passes through the origin indicates homology at the same position in both sequences. Note that no continuous lines are present beyond position -236.

Table 1. *Alu* Repeat Elements in the 5'-Flanking Region of the Human MW Opsin Gene

Position in 5' flanking region	Length of repeat (bp)
-234 to -358	125
-369 to -652	284
-1508 to -1781	273
-2140 to -2427	288

was amplified and cloned. Alignment with the human sequence shows that this fragment contains *Alu* elements at identical positions to the elements present in the human MW gene at positions -234 to -358, and -369 to -652 (Fig. 2).

The homologous region of the human MW and LW upstream sequences show a total of 14 substitutions between the start of the coding sequence and the point of sequence divergence at -237. To determine

Human	CGAAAAGCAACAATTAATACTATTTTCTCATCCACCAGAACGCCAAAAA	-793
DianaG.....A.....	-792
Human	TTAAAAGCCCTAACAAATGTCCAGGGCTGGCGAGAATGTGGCAGAAAGGTGA	-743
Diana	..T.....T.....A.....C.....	-742
Human	TGTCACATACCCTGCAAGTGGGAATCTAAACAGATTCCAGGGTTTGTGTTT	-693
DianaT.....C.....T.....	-692
Human	TTTTTAATCGCAATAGTGGCCCTGTAAATTTTTTTCTGGAGCAGAG	-643
DianaG.....C.....G.....G.....	-647
Human	TTTTGCTCTTCTTGGCCAGGCTGGAGTCCAAATGGCTCGATCTGGCTCAC	-593
DianaA.....	-597
Human	CGCACCTCGAGCTCCAGGTACAAGCGATTCCTCTGTCTCAGCCTCCCA	-543
Diana	T...G.....A...T.....	-547
Human	AGTAGCTGGGAGTACAGGTATTTGCCACTAAGCCCAAGCTAATTTGTTTTT	-493
DianaC.....A...GCA...T.....T.....G.....	-497
Human	ATTTAGTAGAAGCGGGGTTTCACCATGTTAGTCAGGCTGGTGGGAATCTC	-443
DianaT.T.....T.....CA.....	-447
Human	CTGACCTCAGGAGATCTACCCGCTTGGCCTCCCAAGTGTGGGATTAC	-393
DianaT.....C.....A.....T.....	-397
Human	AGGGCTGTGCCACTGTGCCAGCCACTTTTTTTT---AGACAGATCTT	-347
Diana	..T.....TTC.....C.....	-347
Human	GGTCTGTGCCCCAGGCTAGAGTTCAGTGGCCGCATCTCAGCTCACTSCAA	-297
Diana	.T...C.....G...G...F...C...TG.....	-297
Human	CCTCGCCCTCCAGATTCAAGCGATTCTCCTGCCCCGACCTCCAGTAGC	-247
Diana	...T...G.C...C...A...AG.....	-247
Human	TGGGATTACAGGTTTCCAGCAATCCCTCTGAGCCGCCCGGGGGCTCG	-197
DianaA.....	-197
Human	CCTCAGGAGCAAGGAAGCAAGGGGTGGGAGGAGGAGTCTAAGTCCCAGG	-147
DianaC.....	-147
Human	CCCAATTAAGAGATCAGATGGTGTAGGATTTGGGAGCTTTTAAGGTGAAG	-97
Diana	-97
Human	AGGCCCGGGCTGATCCCACTGGCCGGTATAAAGCGCCGTGACCCCTCAGGT	-47
Diana	-47
Human	GACGCACCAGGCGCCGCTGCCGTGGGGACAGGCTTTCCATAGCC	-1
Diana	-1

Figure 2 Alignment of the upstream region of the MW gene in humans and diana monkeys. The *Alu* repeat elements present in both sequences are shaded. Identity between the human and diana sequences is indicated by a dot.

whether these differences are conserved, we have sequenced the region in the chimpanzee *Pan troglodytes* and in two species of cercopithecoid monkeys, the diana monkey *Cercopithecus diana* and the Patas monkey *Erythrocebus patas*. As shown in Figure 3, with the exception of the substitution at the transcription start, all of the differences between the human MW and LW upstream sequences in this region of homology are conserved, although the cercopithecoid MW and LW sequences differ at a few other positions.

The 5'-Flanking Region of the Single MW/LW Opsin Gene in a New World Primate

Using a combination of gene walking and conventional PCR, we have sequenced the entire 5'-flanking region of the marmoset MW/LW opsin gene from the translation start site to the conserved core sequence of the LCR. Figure 4 shows the alignment of the marmoset sequence with the upstream region of the human LW gene. Our amplification strategy involved using a primer designed to the 5' half of the conserved core sequence of the LCR; therefore, the sequence extends only as far as the 3' half of this region. The available sequence from the marmoset is sufficient to conclude, however, that an LCR is present upstream of the single MW/LW gene in a New World monkey.

Sequence identity between the marmoset and hu-

Human LW	GGTCCAGCAAATCCCTCTGAGCCGCCCTTGGGGTCCGCCCTCAGGAGC	-187
Chimp LW	
Diana LW	
Patas LW	
Human MWT.....CC.G.....	
Chimp MWT.....CC.G.....	
Diana MWT.....CC.GA.....	
Patas MWT.....CC.GA.....	
Human LW	AGGGGACCAAGAGGCTGGGAGGAGGCTCAAGTCCAGGCCCAATTAAG	-137
Chimp LW	
Diana LWA.....	
Patas LWA.....	
Human MWA...A...G.....	
Chimp MWA...A...G.....	
Diana MWA...A...G.....C.....	
Patas MWA...A...G.....C.....	
Human LW	AGATCAGGTAGTGTAGGGTTGGGAGCTTTTAAGGTGAAGAGGCCCGGCG	-87
Chimp LW	
Diana LWA.....A.....T...T.....	
Patas LWA.....A.....T...T.....	
Human MWA.G.....A.....	
Chimp MWA.G.....A.....	
Diana MWA.G.....A.....	
Patas MWA.G.....A.....	
	TATA box	
Human LW	TGATCCCAAGGCCAGTATAAAGCGCCGCTGACCCCTCAGGTGATGCGCCAG	-37
Chimp LW	
Diana LWA.....T.A.....	
Patas LWA.....T.A.....	
Human MWT...G.....C...A.....	
Chimp MWT...G.....C...A.....	
Diana MWT...G.....C...A.....	
Patas MWT...G.....C...A.....	
	Codon 1	
Human LW	GGCCGCTGCCGTGGGGACAGGCTTTCCATAGCC ATG	+3
Chimp LW	
Diana LW	..A.T.A.AA.....A.....	
Patas LW	..A.T.A.AA.....A.....	
Human MW	
Chimp MW	
Diana MW	
Patas MW	

Figure 3 Nucleotide sequences of the homologous upstream regions of the MW and LW genes of Old World primates. Identity to the human LW sequence is indicated by a dot.

Duplication of the X-Linked Opsin Genes in the Howler Monkey

A fragment of ~700 bp was amplified that included 620 bp of the 5'-flanking region of the X-linked opsin genes in the howler monkey. The sequencing of nine clones of this amplified fragment showed that they fell into two classes distinguished by four nucleotide differences in the 5'-flanking region at -111, -125, -422, and -431, and a single difference in the coding region at +38 (Fig. 5). This latter substitution will result in an arginine/glutamine amino acid difference between the two howler opsins. The substitution at -111 has been reported previously by Kainz et al. (1998) in a shorter direct-sequenced PCR-amplified fragment from the howler upstream region, but the second substitution at -90 reported by these investigators (these sites are incorrectly numbered in Kainz et al. 1998) was not present in our sequence. Therefore, the presence of two distinct upstream sequences is consistent with the presence of two X-linked opsin genes in this species of New World monkey and we conclude that one is derived from the upstream region of the MW gene and the other from the upstream region of the LW gene. This fragment extends well beyond the 236 bp of homology of the 5' flanking regions of the human MW and LW genes, therefore, indicating that the duplication in the howler monkey differs from that in Old World primates. A 350-bp fragment of the region immediately downstream of the LCR was amplified successfully from howler monkey DNA by using a primer to a conserved region of the marmoset and human LW upstream sequences at position -2896. Cloning and sequencing yielded two distinct sequences that are identical in the LCR core sequence but differ by two

nucleotide substitutions in the less conserved 3' flanking region (Fig. 6). We conclude from this that there are two X-linked LCRs in the howler monkey. Therefore, the duplication in this species must include the entire 5' flanking region of the gene and the LCR. Finally, by using the howler monkey sequence data downstream of the LCR and upstream of the transcription start site to design PCR primers AL⁻ and AL⁺ (Table 2), a fragment of ~2.8 kb that overlapped with both sequences was amplified and cloned. Sequencing of the ends of these clones enabled us to associate, respectively, the upstream sequences 1 and 2 shown in Figure 5 with LCR sequences 1 and 2 shown in Figure 6. The additional sequence that this generated is included in Figure 6.

DISCUSSION

Trichromacy in all Old World primates depends on the presence of two separate X-linked opsin genes that encode a MW and LW opsin, with λ_{\max} values of ~530 and 560 nm, respectively (Bowmaker et al. 1991). The close homology between these two genes indicates that they represent a relatively recent duplication from a single ancestral gene (Hunt et al. 1998) and the presence of the duplication in all Old World primates examined to date (Ibbotson et al. 1992; Deeb et al. 1994; Dulai et al. 1994) places the duplication event at the base of the catarrhine lineage ~40 mya (Goodman et al. 1998). A recent analysis of the coding sequence of these two genes would suggest that the evolution of the spectral shift between the visual pigments encoded by these two genes occurred subsequent to duplication and did not draw on different allelic forms of the polymorphic gene now present in New World primates

Howler 1	AGGAGCTCGACCTCATGGGGGAAAGGGGAAGAAGGGGAGCCGAGAGCAGGATACACGCACGTAGTCAGCCTGCCCGCCCTCACCCGGGGAGGAGCC	-522
Howler 2	
MarmosetA.....T.....A.A.C.....TT.....	-526
Howler 1	CCCTGGCAGAGATCTCGCCTGTTTGGGCCACAGAGATGGTCGGCCTCGAAAGGAGGATCTGGGTCCAAGCCTTGGCCCTGAAAAATCTCTCTATGTTTT	-422
Howler 2C.....C.....	
MarmosetA.CA.....G.....C.G.....	-446
Howler 1	CAAGAGAAAGGAGGATTCGGGCTCTGGGCAGA-TTCATCATCCACCCAA-----GTTGACCTGCCCAAAGGGCGGGCTCTG	-346
Howler 2	
MarmosetT.....T.G.....C.....CCAGTCTGCAATCTTGACCTGG.T.....A.....	-346
Howler 1	GACACTGGAAAGAGAACAGAGGACACGTGAACAGTGAACACAGCGAGGGCCACTGCGGCTCCACAGGCACCTCCACCTGCCCTGGCGGGG--TGGGTGGGG	-248
Howler 2	
MarmosetT.A.....GG.....	-246
Howler 1	CATGTGAGTTTGGTTTCCAGCAAATCCCTCTGAGCCGCCCTCACAGGCTCGCCTCAGGAGCAGGGAAGCAAGGGGTGGGAGGAGTAGGTGTAAGTCCACG	-148
Howler 2	
MarmosetC.....G.....G.....G.....	-148
Howler 1	GCCCAATTAAAGAGATCAGATGGGTAGGGTTTGGGAAGCTTTTAAGGTAAAGAGGCCCGGGCTCATCCACAGGCCACTATAAAGCCCGGTGACGCTCAGG	-48
Howler 2C.....G.....	
MarmosetG.....G.....G.....G.....T.T.....	-48
Howler 1	TTACCAGCCAGGGCTGGCTGCCGTCGGGACAGGGCTTTCATAGCC ATG GGC CAG CAG TGG AGC CTC CAG AGG CTC GCA GGC CGG	+39
Howler 2	
MarmosetG.....	

Figure 5 Nucleotide sequence of the upstream region of the X-linked opsin genes in the howler monkey, aligned with the equivalent region of the marmoset gene. Identity of howler sequence 2 and marmoset to howler sequence 1 is indicated by a dot. Sequence deletions are indicated by dashes. Howler monkey sequence, GenBank accession nos. AF156715 and AF156716.

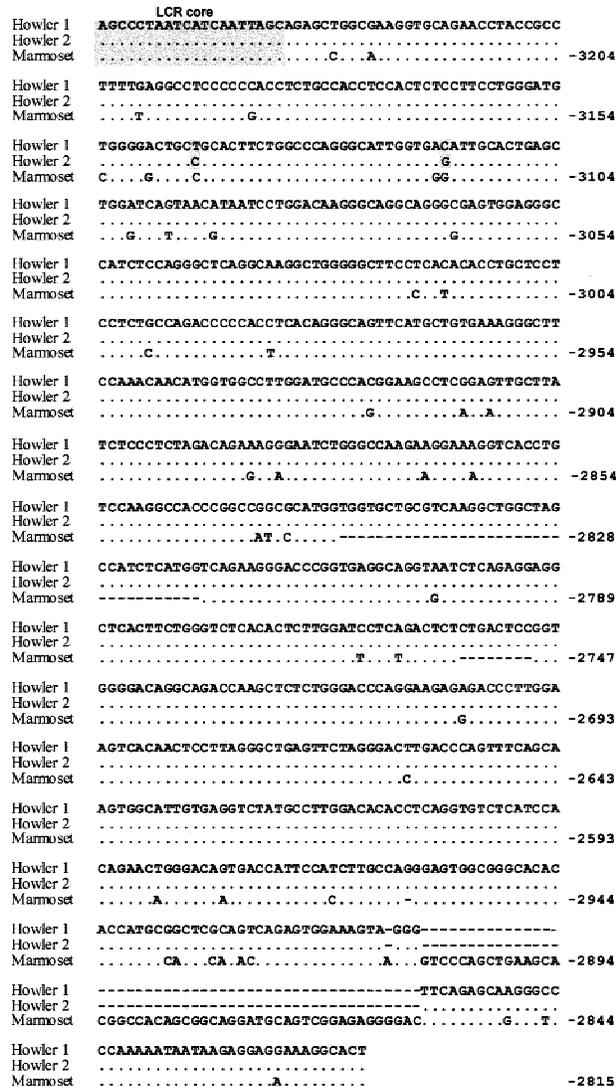


Figure 6 Nucleotide sequence of the two locus control regions identified in the howler monkey, aligned with the corresponding region in the marmoset. The numbering of the marmoset sequence corresponds to the full sequence shown in Fig. 3. Identity of the other two sequences to howler sequence 1 is indicated by a dot.

(Hunt et al. 1998). The single X-linked opsin gene that is present in most New World monkeys is clearly orthologous to both genes in Old World species (Ibbotson et al. 1992; Williams et al. 1992; Dulai et al. 1994; Hunt et al. 1998) and therefore, most likely represents the ancestral organization before gene duplication.

Previous sequence analysis of the 5' flanking regions of the human MW and LW genes has demonstrated that the two regions are homologous for only 236 bp immediately upstream of the translation start site (Nathans et al. 1986). This raises the question of the origin of the different sequences beyond this region. By sequencing the 5' flanking region of the mar-

moset, a New World monkey with a single X-linked opsin gene and a polymorphism-based trichromacy, we have been able to demonstrate that this region shows extensive homology to the human LW gene but not to the human MW gene. In fact, the homology between the marmoset and the human MW sequences is limited to the same 236 bp immediately upstream of the translation start as found for the two human sequences. From this, we can deduce that the duplicated opsin gene in Old World primates includes only this short upstream region, and that the 5' flanking region of the human LW gene represents the ancestral sequence.

Recent work by Hanna et al. (1997) has demonstrated that, in humans, a novel gene called *TEX28*, which is expressed only in the testes, lies at the 3' end of the opsin gene array on the X chromosome. A truncated copy of this gene is also present immediately downstream of the opsin gene closest to the LCR, and any additional copies of MW genes are also flanked by truncated copies of *TEX28*. It is only the final opsin gene at the 3' end of the array that is flanked by a complete copy of *TEX28*. Therefore, the duplication in Old World primates encompasses a 35-kb region that includes a complete copy of the opsin-coding region, 236 bp of the upstream flanking sequence, and a truncated copy of the *TEX28* gene. With the exception of the opsin gene adjacent to the LCR, all other genes within the array now possess a minimal promoter region of ~190 bp of the original flanking region upstream of the transcription start site. This duplicated region was inserted either just upstream of the original opsin gene or into the first intron of the *TEX28* gene, as depicted in Figure 7. If the former alternative is correct, then the duplication must have occurred by an unequal crossover mechanism that avoided any sequence duplication or deletion at the point of insertion, as there is no evidence for any indels in the marmoset and human LW sequences in the immediate vicinity of the insertion. The preservation of the exact sequence might have been less important if the insertion was into the intron of the *TEX28* gene. It has been suggested that the LW opsin was most probably ancestral in primates (Nei et al. 1997). This does not, however, distinguish between the above two alternatives as, subsequent to duplication, evolution of a MW sequence could have occurred in either gene and it does not follow that the present LW gene was the ancestral sequence.

A striking feature of the upstream region of the human MW gene beyond the region of homology is the presence of a number of *Alu* repeat elements, and at least the first two of these are also present in identical positions in the upstream region of the MW gene of the diana monkey. This would indicate that the juxtapositioning of these elements with the MW opsin gene

Table 2. Sequence and Position of PCR Primers

Primer	Position	Sequence (5' to 3')
UNI-33		TTTTTTTTTTTTTTGTTTGTGTGGGGGGTT
UNI-17		TTTTGTTTGTGTGGG
Op1 ⁻	+94	TGCTGGACTGGGTGCTGTCTCATAGC
Op2 ⁻	+78	TCATAGCTGTCTCGGGATGGCG
Op1 ⁺	+254	CCAGTAGCTGGGATTACAGG
Op2 ⁺	-151	TGGGAGGAGGAGGCTAAGTCC
Up(A ⁻)	-130	GCTCCCAAATCCTACACCATC
Up(B ⁻)	-233	AGAGGGATTGCTGGAAC
Up(C ⁻)	-2350	AGACAGTAGACAGGTGGTGCC
MW ⁻	-114	AAATCCTACACCATCTGATCTC
MW ⁺	-865	AGGTGATCCGGGAATAGTAAG
NWMLCR ⁻	-360	CTGTCCAGTGCCCGCCTTGG
MarUp(A ⁺)	-627	GAAGTAGGTGCCAGCTCATGG
MarLCR(C ⁺)	-2896	CCTCAGAATTGCTTATCTCCCTC
MarLCR(D ⁻)	-1015	TGCCTCCTTTTGGGTTTGTTC
MarLCR(D ⁺)	-2221	TTTGTATTCAATCATCCG
MarLCR(E ⁻)	-1257	GTGCTAGGTGCCAAGCTAGG
MarLCR(E ⁺)	-1900	CGATGGACACAGTTTGCTTCC
LCR(C ⁺)	-3484	GACTTGATCTTCTGTTAGCCCTAATC
AL ⁻	-380	GGATGATGAATCTGCCAGAG
AL ⁺	-3225	CGAAGGTGCAGAACCTACCG

Except for Op1⁺, Op1⁻ and Op2⁻, which are based on the human MW and LW sequences, and AL⁻, which is based on the howler monkey sequence, the nucleotide numbering system refers to the marmoset X-linked opsin gene, where +1 would be the first base of the coding sequence.

occurred at the same time as the gene duplication rather than as a result of a more recent insertion. The first of these elements, a half *Alu*, is found at the point of insertion of the original duplication. Because *Alu* repeat elements have been implicated in unequal crossing over within other genes (e.g., see Campbell et al. 1995; Rudiger et al. 1995; Hartevelde et al. 1997; Levran et al. 1998) and in the insertion event that resulted in the "long" intron 1 of the human LW opsin gene (Meagher et al. 1996), it is possible that the presence of *Alu* sequences was a contributory factor in the generation of the original opsin gene duplication, and in the generation of additional copies of the MW gene (Drummond-Borg et al. 1989; Ibbotson et al. 1992). By the same mechanism, these sequences may also be important in promoting unequal crossing over within the opsin gene array that leads to either the complete loss of opsin genes or the generation of hybrids. Such changes are considered to be responsible for anomalous trichromacy and dichromacy in humans (Nathans et al. 1986; Deeb et al. 1992).

The close proximity of the duplication insertion breakpoint to the transcription start site of the MW gene would imply that the promoter for this gene is contained within the remaining 199 bp of upstream sequence. This has been confirmed by an *in vitro* study of the activity of this region in driving the expression of the human MW and LW genes in transfected WERI Rb1 cells, a retinoblastoma cell line that is known to express cone opsins (Shaaban and Deeb 1998). The

minimal promoter, numbered from immediately upstream of codon one, was defined in this study as the region from -64 to -1, with positive elements within the region from -83 to -171, and negative elements from -171 to -240. This study was unable to provide, however, any information on the mechanism of selective gene expression that ensures that only one visual pigment is produced in a single cone photoreceptor, although differences in the activity of the two promoters were clearly apparent. Given the divergence of both promoters from that of the marmoset, it is tempting to speculate that the control of differential gene expression in Old World primates may also reside in this region, and our finding that all the upstream differences, except that at position +1 within

this minimal promoter region in the human and chimpanzee MW and LW genes, are conserved in two species of cercopithecoid monkeys is supportive of this interpretation.

The presence of an LCR upstream of the single X-linked opsin gene in the marmoset is consistent with the idea that this highly conserved region is required for the activation of gene transcription in this chromosomal region. It has been suggested that the LCR interacts by a looping mechanism with the adjacent gene promoter to form a transcriptionally active chromatin domain (Festenstein et al. 1996; Milot et al. 1996) through protein-protein interactions of transcription factors that bind to both the promoter and LCR (Shaaban and Deeb 1998). In this case, sequence differences in the MW and LW opsin promoters may determine specificity of gene expression through the differential binding of such transcription factors. However, this would require a further level of gene regulation and the alternative possibility that the interaction of the LCR with one or other promoter region is mutually exclusive, provides a simpler explanation. The presence of an LCR in New World monkeys argues, however, that the primary role of this region is not the selection of either a MW or LW opsin gene for expression as only a single gene is present in most New World primates; the presence of the LCR must therefore predate the Old World opsin gene duplication.

The howler monkey is unique among New World primates in possessing separate MW and LW genes.

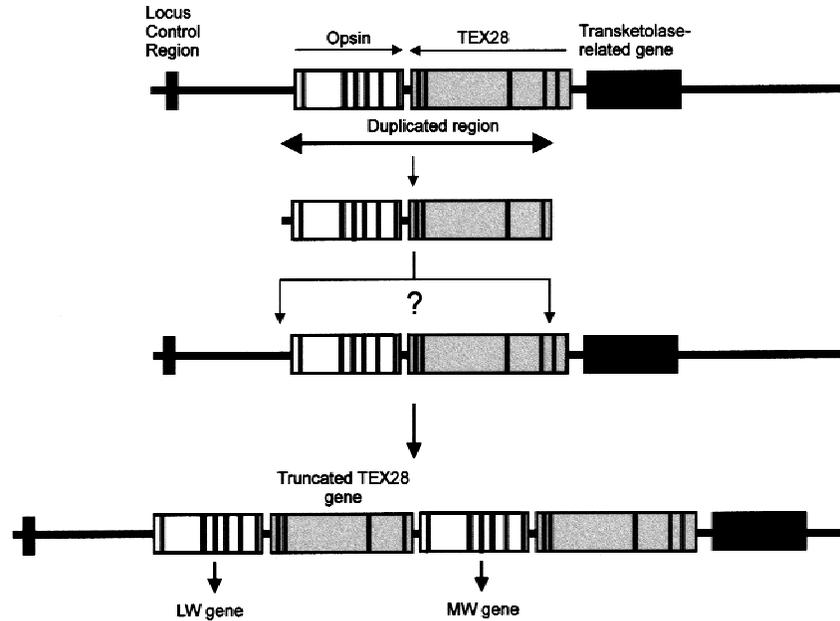


Figure 7 Origin of the opsin gene duplication on the X chromosome of Old World primates. The opsin and *TEX28* genes are shown as open boxes showing the approximate positions of exons and introns. The LCR and transketolase-related gene are shown as filled boxes. Details of the *TEX28* and transketolase-related genes are from Hanna et al. (1997).

Therefore, in this respect, it resembles the Old World primates. The presence of flanking regions to the two X-linked opsin genes in this species that differ at a small number of sites was first reported by Kainz et al. (1998) and the more extensive sequence analysis of this region presented here confirms their observation. The extensive identity within the 620-bp region immediately upstream of the howler genes contrasts with the divergence of the upstream region within 237 bp of the translation start site for the human MW and LW genes. The howler monkey and Old World primate duplications are, therefore, quite distinct and the homology in this region between the two howler genes, with only four single nucleotide differences in >600 bp of DNA, indicates that the howler monkey duplication is much more recent than that in Old World primates. In addition, we have been able to clone and sequence two distinct LCRs from the DNA of a single male howler monkey that differ by two nucleotide substitutions in the 3' flanking region of the conserved core sequence of the LCR, and to link these two LCR sequences to the two sequences of the immediate upstream region, thus defining two complete 5' flanking regions of the howler opsin genes. Therefore, it would appear that the duplication in the howler monkey involved not only the complete coding region of the opsin gene but also the entire upstream region, including the LCR.

In typical New World monkeys with a single polymorphic X-linked opsin gene, trichromacy is achieved in heterozygous females by random X inactivation. In

contrast, separate MW and LW opsin genes will generate trichromacy only if a mechanism exists to limit expression to only one gene per cone photoreceptor. The presence of two LCRs in the howler monkey would indicate that the mechanism of selective opsin gene expression may differ from that in Old World primates. Whether the limited sequence differences within the promoter regions and LCRs of these two genes in howler monkeys are sufficient for this process remains to be determined.

METHODS

DNA Samples

DNA was isolated from a blood sample taken from a male black howler monkey *Alouatta seniculus*. DNA was isolated from liver tissue of a male marmoset *Callithrix jacchus*, with the P563 allele of the X-linked opsin gene, using a standard phenol-chloroform method as described previously (Williams et al. 1992). DNA from the two cercopithecoid monkeys, the diana monkey *Cercopithecus diana* and the patas monkey *Erythrocebus patas*, was isolated as described by Ibbotson et al. (1992), and from the chimpanzee as described by Dulai et al. (1994).

Amplification of the 5'-Flanking Regions of Opsin Genes

The sequences of all primers used in PCR amplifications are listed in Table 2.

Human and Diana Monkey MW Gene

Three clones containing the upstream region of the human MW gene were isolated from a human cosmid library (Cachon-Gonzales 1991), using exon 1 of the human MW gene as a probe. The presence of the MW coding region was confirmed by PCR amplification of exon 2. One of these clones was prepared for sequencing by digestion with *EcoRI*, followed by the ligation of *EcoRI* Vectorette 1 linkers. The Vectorette 1 primer was then used in combination with a gene-specific primer (Op1⁻) to generate a fragment of ~2.3 kb. This was direct sequenced using Op1⁻ as a sequencing primer. The sequenced region was extended in two more steps to -1185 through the use of two additional sequencing primers designed from the sequenced region. Approximately 750 bp of the upstream region of the diana monkey MW gene was amplified using primer pair MW⁻ and MW⁺. The resulting fragment was cloned and sequenced.

Marmoset MW/LW Gene

A combination of conventional PCR and gene walking was used to amplify the upstream region of the marmoset gene. Gene walking allows unknown flanking regions to be amplified from a known sequence. Initially, a 245-bp fragment that included the first exon and extended upstream to -151 bp

was amplified with primers Op2⁺ and Op1⁻, designed from the X-linked human opsin genes (Nathans et al. 1986). This sequence was then extended upstream to -205 bp, using an unpredictably primed PCR developed by Domínguez and López-Larrea (1994) and modified by Bellingham et al. (1998). Gene-specific nested primers used in the gene walk were Op1⁻, Op2⁻, and the nested "universal" walking primers were UNI-33 and UNI-17. The first round UP-PCR-1 mix contained 10 ng of DNA, 12.5 pmoles of primer UNI-33, 0.5 nmole each of dATP, dCTP, dGTP, and dTTP, 3 mM MgCl₂, and 5 µl of 10× NH₄ buffer in a total volume of 23 µl. The mix was heated at 94°C for 60 sec to denature the DNA, during which 0.25 units of *Taq* polymerase was added in a volume of 1 µl. The incubation temperature was then lowered to 80°C for 30 sec, then rapidly to 15°C for 120 sec, and finally held at 25°C for 10 min. The sample tubes were then placed in a thermal cycler, initially at 72°C for 60 sec before DNA denaturation at 94°C for 60 sec during which 12 pmoles of the outer Op1⁻ primer was added in a volume of 1 µl. The sample was then heated at 94°C for 1 sec, 62°C for 30 sec, and 72°C for 30 sec for 35 cycles, followed by a final extension at 72°C for 5 min. The UP-PCR-2 contained 1 µl of 1:1000 (vol/vol) dilution of the UP-PCR-1, 12.5 pmoles each of primers Op2⁻ and UNI-17, 0.5 nmole each of dATP, dCTP, dGTP, and dTTP, 3 mM MgCl₂, and 5 µl of 10× NH₄ buffer in a total volume of 24 µl. *Taq* polymerase (0.25 units) in 1 µl was added during the DNA denaturation step, as done previously, and the sample cycled at 94°C for 1 sec, 56°C for 1 sec, and 72°C for 30 sec for 35 cycles, followed by a final extension at 72°C for 5 min.

A further round of UP-PCR with a new set of primers [Up(A-) and Up(B-)] together with UNI-33 and UNI-17 generated a fragment of ~550 bp, which extended upstream to base -659. On the basis of the sequence obtained from this fragment, further primers [NWM LCR⁻ and MarUp(A⁺)] were designed to allow amplification both upstream and downstream back into the gene. Primer NWM LCR⁻ was combined with primer LCR(C⁺) to amplify a fragment of ~3000 bp, which extended from the LCR core to a region in the proximal promoter. This fragment was fully sequenced using a set of nested primers [MarLCR(C⁻), MarLCR(D⁺), MarLCR(E⁺), MarLCR(D⁻), and MarLCR(E⁻)].

The Conserved Region of the Chimpanzee, Diana Monkey, and Patas Monkey MW and LW Genes

The sequence of the LW gene of the chimpanzee, diana and patas monkeys was obtained by gene walking using Op1⁻ and Op2⁻ as gene-specific primers. This generated fragments of ~400 bp that were both direct sequenced using Op2⁻ and UNI-17 as sequencing primers, and sequenced after cloning. The MW sequences were obtained by conventional PCR using reverse primer Op2⁻ and forward primer Op1⁺ that sits at the boundary of the region of homology between the MW and LW genes. The fragments were both directly sequenced and sequenced after cloning.

Howler MW and LW Genes

The upstream regions were initially amplified as two discrete blocks by conventional PCR using primer pairs MarUp(A⁺) with Op2⁻, and LCR(C⁺) with Up(C⁻), using standard PCR parameters with annealing temperatures of 61°C and 57°C, respectively. Fragments of ~700 bp and 900 bp, respectively, were generated by these amplifications. Each fragment was both directly sequenced and sequenced after cloning. A

primer pair (AL⁻ and AL⁺) was designed from these two sequences that would amplify the intervening region of ~2.8 kb. The resulting fragment was cloned and a ~350-bp region at each end was sequenced.

DNA Cloning and Sequencing

All amplified fragments were sequenced directly on an Applied Biosystems model 373a automated sequencer. Before cycle sequencing, the PCR mix was cleaned with a Centricon (Princeton, Inc.) column. Approximately 100 ng of DNA was combined with 4 µl of ABI PRISM Dye Terminator and 3.5 pmoles of an appropriate sequencing primer, and cycled on a Perkin Elmer 9600 PCR thermal cycler. Selected PCR fragments were cloned into pTag or pGEM-T-Easy, before cycle sequencing. Alignments were performed using GeneWorks.

ACKNOWLEDGMENTS

This work was supported by a grant from the Medical Research Council. We are grateful to Dr. C. Julliot for the sample of howler monkey DNA.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Bellingham, J., A.G. Morris, and D.M. Hunt. 1998. The rhodopsin gene of the cuttlefish, *Sepia officinalis*: cDNA sequence, gene structure and spectral tuning. *J. Exp. Biol.* **201**: 2299–2306.
- Bowmaker, J.K., S. Astell, D.M. Hunt, and J.D. Mollon. 1991. Photosensitive and photostable pigments in the retinæ of Old World monkeys. *J. Exp. Biol.* **156**: 1–19.
- Cachon-Gonzalez, B. 1991. "Linkage analysis in familial adenomatous polyposis families in the United Kingdom, and a search for highly polymorphic markers." Ph. D. thesis, University of London, London, UK.
- Campbell, C., I. Marondel, K. Montgomery, K. Krauter, and R. Kucherlapati. 1995. Unequal homologous recombination of human DNA on a yeast artificial chromosome. *Nucleic Acids Res.* **23**: 3691–3695.
- Deeb, S.S., D.T. Lindsey, Y. Hibiya, E. Sanocki, J. Winderickx, D.Y. Teller, and A.G. Motulsky. 1992. Genotype-phenotype relationships in human red/green color-vision defects: Molecular and psychophysical studies. *Am. J. Hum. Genet.* **51**: 687–700.
- Deeb, S.S., A.L. Jorgensen, L. Battisti, L. Iwasaki, and A.G. Motulsky. 1994. Sequence divergence of the red and green visual pigments in great apes and humans. *Proc. Natl. Acad. Sci.* **91**: 7262–7266.
- Domínguez, O. and C. López-Larrea. 1994. Gene walking by unpredictably primed PCR. *Nucleic Acids Res.* **22**: 3247–3248.
- Drummond-Borg, M., S.S. Deeb, and A.G. Motulsky. 1989. Molecular patterns of X chromosome-linked color vision genes among 134 men of European ancestry. *Proc. Natl. Acad. Sci.* **86**: 983–987.
- Dulai, K.S. 1996. "Molecular genetics of the middlewave and longwave sensitive opsin genes of higher primates." Ph.D. thesis, University of London, London, UK.
- Dulai, K.S., J.K. Bowmaker, J.D. Mollon, and D.M. Hunt. 1994. Sequence divergence, polymorphism and evolution of the middle-wave and long-wave visual pigment genes of Great Apes and Old World monkeys. *Vision Res.* **34**: 2483–2491.
- Feil, R., G. Palmieri, M. d'Urso, R. Heilig, I. Oberle, and J.L. Mandel. 1990. Physical and genetic mapping of polymorphic loci in Xq28 (DXS15, DXS52, and DXS134): Analysis of a cosmid clone and a yeast artificial chromosome. *Am. J. Hum. Genet.* **46**: 720–728.
- Festenstein, R., M. Tolaini, P. Corbella, C. Mamalaki, J. Parrington, M. Fox, A. Miliou, M. Jones, and D. Kiuoussis. 1996. Locus control region function and heterochromatin-induced position effect variegation. *Science* **271**: 1123–1125.
- Goodman, M., C.A. Porter, J. Czelusniak, S.L. Page, H. Schneider, J.

- Shoshani, G. Gunnell, and C.P. Groves. 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* **9**: 585–598.
- Hanna, M.C., J.T. Platts, and E.F. Kirkness. 1997. Identification of a gene within the tandem array of red and green color pigment genes. *Genomics* **43**: 384–386.
- Harteveld, K.L., M. Losekoot, R. Fodde, P.C. Giordano, and L.F. Bernini. 1997. The involvement of Alu repeats in recombination events at the alpha-globin gene cluster: characterization of two alpha zero-thalassaemia deletion breakpoints. *Hum Genet* **99**: 528–534.
- Hunt, D.M., K.S. Dulai, J.A. Cowing, C. Julliot, J.D. Mollon, J.K. Bowmaker, W.-H. Li, and D. Hewett-Emmett. 1998. Molecular evolution of trichromacy in primates. *Vision Res.* **38**: 3299–3306.
- Ibbotson, R.E., D.M. Hunt, J.K. Bowmaker, and J.D. Mollon. 1992. Sequence divergence and copy number of the middle- and long-wave photopigment genes in Old World monkeys. *Proc. Roy. Soc. Lond.* **B247**: 145–154.
- Jacobs, G.H., M. Neitz, and J. Neitz. 1996a. Mutations in S-cone pigment genes and the absence of color vision in two species of nocturnal primate. *Proc. Roy. Soc. Lond.* **B263**: 705–710.
- Jacobs, G.H., M. Neitz, J.F. Deegan, and J. Neitz. 1996b. Trichromatic color vision in New World monkeys. *Nature* **382**: 156–158.
- Kainz, P.M., J. Neitz, and M. Neitz. 1998. Recent evolution of uniform trichromacy in a New World monkey. *Vision Res.* **38**: 3315–3320.
- Levrin, O., N.A. Doggett, and A.D. Auerbach. 1998. Identification of Alu-mediated deletions in the Fanconi anemia gene FAA. *Hum Mutat* **12**: 14–152.
- Meagher, M.J., A.L. Jorgensen, and S.S. Deeb. 1996. Sequence and evolutionary history of the length polymorphism in intron 1 of the human red photopigment gene. *J. Mol. Evol.* **43**: 622–630.
- Milot, E., J. Strouboulis, T. Trimborn, M. Wijgerde, E. de Boer, A. Langeveld, K. Tan-Un, W. Vergeer, N. Yannoutsos, F. Grosveld, and P. Fraser. 1996. Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* **87**: 105–114.
- Mollon, J.D., J.K. Bowmaker, and G.H. Jacobs. 1984. Variations of color vision in a New World primate can be explained by a polymorphism of retinal photopigments. *Proc. Roy. Soc. Lond.* **B222**: 373–399.
- Nathans, J., D. Thomas, and D.S. Hogness. 1986. Molecular genetics of human color vision: The genes encoding blue, green and red pigments. *Science* **232**: 193–203.
- Nathans, J., C.M. Davenport, I.H. Maumenee, R.A. Lewis, J.F. Hejtmancik, M. Litt, E. Lovrien, R. Weleber, B. Bachynski, F. Zwas, et al. 1989. Molecular genetics of human blue cone monochromacy. *Science* **245**: 831–838.
- Nei, M., J. Zhang, and S. Yokoyama. 1997. Color vision of ancestral organisms of higher primates. *Mol. Biol. Evol.* **14**: 611–618.
- Rudiger, N.S., N. Gregersen, and M.C. Kielland-Brandt. 1995. One short well conserved region of Alu-sequences is involved in human gene rearrangements and has homology with prokaryotic chi. *Nucleic Acids Res* **23**: 25–260.
- Shaaban, S.A. and S.S. Deeb. 1998. Functional analysis of the promoters of the human red and green visual pigment genes. *Invest. Ophthalm. Vis. Sci.* **39**: 885–896.
- Shyue, S.-K., D. Hewett-Emmett, H.G. Sperling, D.M. Hunt, J.K. Bowmaker, J.D. Mollon, and W.-H. Li. 1995. Adaptive evolution of color vision genes in higher primates. *Science* **269**: 1265–1267.
- Vollrath, D., J. Nathans, and R.W. Davis. 1988. Tandem array of human visual pigment genes at Xq28. *Science* **240**: 1669–1672.
- Wang, Y., J.P. Macke, S.L. Merbs, D.J. Zack, B. Klaunberg, J. Bennet, J. Gearhart, and J. Nathans. 1992. A locus control region adjacent to the human red and green visual pigment genes. *Neuron* **9**: 429–440.
- Williams, A.J., D.M. Hunt, J.K. Bowmaker, and J.D. Mollon. 1992. The polymorphic photopigments of the marmoset: Spectral tuning and genetic basis. *EMBO J.* **11**: 2039–2045.
- Zhao, Z., D. Hewett-Emmett, and W.H.J. Li. 1998. Frequent gene conversion between human red and green opsin genes. *Mol. Evol.* **46**: 494–496.

Received November 10, 1998; accepted in revised form May 25, 1999.