

Nucleotide polymorphisms upstream of the X-chromosome opsin gene array tune L:M cone ratio

KAREN L. GUNTHER,¹ JAY NEITZ,² AND MAUREEN NEITZ²

¹Department of Psychology, Wabash College, Crawfordsville, Indiana

²Department of Ophthalmology, Medical College of Wisconsin, Milwaukee, Wisconsin

(RECEIVED October 8, 2007; ACCEPTED January 31, 2008)

Abstract

In support of the long-held idea that cone ratio is genetically determined by variation linked to the X-chromosome opsin gene locus, the present study identified nucleotide differences in DNA segments containing regulatory regions of the L and M opsin genes that are associated with significant differences in the relative number of L versus M cones. Specific haplotypes (combinations of genetic differences) were identified that correlated with high versus low L:M cone ratio. These findings are consistent with the biological principle that DNA sequence variations affect binding affinities for protein components of complexes that influence the relative probability that an L versus M opsin gene will be silenced during development, and in turn, produce variation in the proportion of L to M cones.

Keywords: Cone photoreceptor, Color vision, X-chromosome, Cone opsin, Genes

Introduction

The relative numbers of long-wavelength (L) to middle wavelength sensitive cones is highly variable among men with normal color vision (DeVries, 1946; Rushton & Baker, 1964; Dartnall et al., 1983; Pokorny et al., 1991; Roorda & Williams, 1999; Carroll et al., 2002; Hofer et al., 2005) and it has long been suspected that variation in L:M cone ratio is X-linked (DeVries, 1946; Smallwood et al., 2002; McMahon et al., 2008). L and M photo receptors represent a cell type distinct from S cones (Bumsted et al., 1997), and evidence has accumulated recently indicating that L and M cones represent essentially a single cell type differentiated only by which X-chromosome opsin gene is selected for expression (Smallwood et al., 2002). In humans and other Old World primates, the selection mechanism has been proposed to involve a competition between activation, and silencing of adjacent cone opsin genes (Knoblauch et al., 2006). During development, all but one of the opsin genes are silenced so that mature adult L and M cones express a single opsin gene (Hagstrom et al., 2000). In a previous study, the flicker photometric electroretinogram was used to estimate the percentage of cones that are L (defined as $100 * L / (L + M)$) in a sample of males with normal color vision (Carroll et al., 2002), and the distribution of the percentage of L cones passed a statistical test for normality (Kolmogorov-Smirnov Test). A characteristic of inherited quantitative traits such as the percentage of L cones is that the greater the number of genetic loci that contribute to the phenotype, the more closely the distribution of the

trait will approach a normal distribution (Fisher, 1918; Strachan & Read, 2000). The distribution of percent L cones thus is consistent with the cumulative effects of multiple genetic alterations contributing to the relative proportion of L to M cones. Polymorphisms at the X-chromosome cone opsin locus emerge as the most likely candidates for causing variation in the relative numbers of L and M cones by virtue of their potential to influence the relative strengths of activation versus silencing of the opsin genes and their ability to cumulatively produce a relatively broad, continuous range of proportions of L to M cones. We tested the hypothesis that genetic variation in the ~8000 base pair region adjacent to the protein coding sequence for the L opsin gene on the X-chromosome affects the relative proportion of L to M cones.

Materials and methods

Subjects

All experiments involving human subjects were conducted in accordance with the Declaration of Helsinki, and were approved by the Institutional Review Board at the Medical College of Wisconsin. Subjects were 94 males who had previously passed several standard tests for color vision deficiency including Ishihara's 24-plate test, Richmond HRR 2002 edition, and the Neitz Test of Color Vision. Subjects identified themselves as being either of Caucasian descent ($n = 67$) or of African descent ($n = 27$). The proportion of L cones, expressed as the percentage of L plus M cones that were L, had been determined previously for each subject from a combination of genetics and the flicker photometric electroretinogram (FP-ERG) (Carroll et al., 2002; McMahon et al., 2008). Here we applied a correction factor to the estimated pro-

Address correspondence and reprint requests to: Maureen Neitz, Department of Ophthalmology, Medical College of Wisconsin, 925 North 87th Street, Milwaukee, WI 53226-4812. E-mail: mneitz@mcw.edu

portion of L cones to account for the 1.5 times greater contribution to the FP-ERG signal for each M cone compared to each L cone (Hofer et al., 2005).

DNA isolation and the polymerase chain reaction

Nucleic acid was isolated from buccal swabs or whole blood using PureGene DNA extraction kits. A series of polymerase chain reactions (PCRs) were designed to provide full coverage of the genomic DNA region upstream of the L opsin gene (amplicons 1–31, Table 1) and for the region upstream of the M opsin gene (amplicon 32, Table 1). X-chromosome nucleotides 152,926,644 through 152,930,676 (amplicons 13–31, Table 1) and 152,968,633 through 152,969,085 (amplicon 32) were sequenced for all 94 subjects (nucleotide numbers correspond to the May 2004 assembly of the Human Genome) as described below under DNA sequencing. The full reference sequence of the region examined can be accessed at <http://genome.ucsc.edu>. Nucleotides 152,923,038 through 152,926,801 (amplicons 1–13, Table 1) were analyzed for a subset of subjects, most of whom had a percentage of L cones that deviated from the mean for their ethnic group by more than one standard deviation, and thus represent the extremes for high and low percentage of L cones in our subject pool. For reference, the L opsin gene coding sequence starts at nucleotide 152,930,605 (note: amplicon 31 extends into the L opsin protein-coding region), and the M opsin coding sequence starts at nucleotide 152,969,104.

As indicated in Table 1, the polymerase was either AmpliTaq Gold (0.25 μ l), rTth DNA Polymerase XL (1 μ l), Takara (0.5 μ l), or a mixture of AmpliTaq Gold and Pfu in a 9:1 volume: volume ratio (0.25 μ l) (Gunther et al., 2006). Final reaction volumes were 100 μ l for reactions with rTth XL enzyme and 50 μ l for all other reactions. Final concentrations for standard reaction conditions were 200 μ M of each of dATP, dTTP, dCTP, and dGTP, 1.4 mM either MgCl₂ (AmpliTaq Gold) or Mg(OAc)₂ (rTth XL), 0.25 μ l of AmpliTaq Gold or 1 μ l of rTth XL enzyme, and 1 μ l genomic DNA. In order to improve the yield of PCR product, we used non-standard conditions for amplicons 20, 21, 24, and 29 (Table 1). The final concentration of dATP, dTTP, dCTP, and dGTP was 400 mM for amplicon number 20 and 100 mM for amplicons 21, 24, and 29. The final concentration of MgCl₂ was 1.0 mM for amplicons 21, 22, 23, and 29, and 3 mM for amplicon 24. For amplicon number 20, the Takara enzyme was used with the manufacturer-provided buffer containing MgCl₂.

Standard thermal cycling conditions for reactions using the AmpliTaq Gold included an initial 9 min incubation at 95°C to activate the enzyme, followed by 37 cycles of 45 s at 94°C, 45 s at the annealing temperature given in Table 1, and 1 min at 72°C. A final incubation was done for 7 min at 72°C. Standard thermal cycling conditions for reactions using the rTth XL enzyme included an initial incubation for 1 min at 94°C, followed by 35 cycles of 15 s at 94°C, 30 s at the annealing temperature given in Table 1, and 1 min at 72°C. Reactions were then incubated for 10 min at 72°C. Standard thermal cycling conditions for reactions using AmpliTaq Gold mixed with Pfu included an initial 10 min incubation at 94°C followed by 34 cycles of 20 s at 94°C, 1 min at a changing annealing temperature, and 1 min at 72°C. In the first cycle, the annealing temperature was 2°C above the annealing temperature given in Table 1, and it was decreased by 0.5°C for each successive cycle through cycle 14. Cycles 15 through 34 used the same temperature as in cycle 14 and this was 5°C below the temperature given in Table 1. After cycle 34, reactions were incubated for 10 min at 72°C.

dHPLC analysis

Amplicons 14, 15, 16, 17, 19, 22, 26, 28, and 32 were screened for sequence variations by analyzing heteroduplex formation in dHPLC on the Wave-MD (Transgenomic, Omaha, NE). Heteroduplexes form when two strands of DNA differ by as little as a single nucleotide; however, heteroduplexes should not form in the samples from individual male subjects because they should have one X-chromosome. To screen for DNA polymorphisms by heteroduplex formation, an amplicon from each subject was mixed with the same amplicon from control subject 015, and in a separate tube, with the same amplicon from control subject 005. The nucleotide sequences of amplicons from control subjects were determined directly, as described below. To facilitate heteroduplex formation, amplicon mixtures were incubated in a thermal cycle with an initial temperature of 95°C, then decreased to 48°C in 0.5°C increments at 20-s intervals, then held at room temperature until assayed. Heteroduplex analysis was performed on the Wave-MD according to the manufacturer's instructions, and using one or two temperatures. Two temperatures were used for amplicons lacking known polymorphic positions in order to increase our chances of detecting polymorphisms. The amplicons and temperatures used for dHPLC analysis were as follows: amplicon 14 64.7°C and 64.8°C, amplicon 15 61.5°C and 61.8°C, amplicon 16 63.0°C and 63.3°C, amplicon 17 64.5°C and 64.7°C, amplicon 19 61.6°C and 61.8°C, amplicon 22 58.4, amplicon 26 61.6°C and 61.9°C, amplicon 28 65.2°C and amplicon 32 62.5°C. DNA polymorphisms discovered by this method were verified by direct sequencing for all subjects who had the low frequency variant and for a subset of subjects with the high frequency variant.

DNA sequencing

Screening for DNA sequence polymorphisms was done by direct sequencing of the two dHPLC control subjects (015 and 005), for spot-checking dHPLC results, and for all amplicons in Table 1 except for those listed above for which dHPLC screening was performed. Amplicons were sequenced with the primers indicated in Table 1 that were used in the initial PCR amplification. Sequencing reactions were done using AmpliTaq FS kit (Applied Biosystems, Foster City, CA) following the manufacturer's recommendations and analyzed on an ABI 3100 XL or an ABI 3100 Avant.

Fragment analysis

Amplicons 2 and 7 (Table 1) each contained homopolymeric regions (extended regions of a single repeated nucleotide) which could not be determined unambiguously by direct sequencing. Fragment analysis was used instead to evaluate the homopolymeric regions for variations in length across subjects. Amplicons 2 and 7 were amplified as described above except that one primer in each pair was 5'-end labeled with the fluorescent dye 6FAM and PCR-amplified (Table 1). PCR products were purified using Microcon YM-30 Centrifugal Filter Devices (Millipore), diluted 1:2, 1:10, and 1:100 with water, and 0.5 μ l of the diluted PCR product was mixed with 9.25 μ l formamide and 0.25 μ l GeneScan LIZ500 size standard. Mixtures were heat denatured for 4 min at 95°C, cooled on ice for 5 min, run on an ABI 3130XL and analyzed using GeneMapper software.

Results

Twenty-eight single nucleotide polymorphisms (SNPs) and four insertion/deletions (indels) were identified within the 7,567 base

Table 1. Amplicons and primers

#	Region amplified	T _A °C	Enzyme	Forward primer (concentration (nM)) ¹	Reverse primer (concentration (nM))
1	152923038–152923625	63	rTth XL	<u>cgaccaggatccaccctt</u> tc (900)	gcggaaggagagagacataatgg (900)
2	152923498–152923661	61	rTth XL	cagccttgctccggattc (300)	ttcaaattcacttcatgttcct-6FAM (300)
3	152923498–152924077	61	AmpliTaq Gold	cagccttgctccggattc (500)	cctgctgggtctgtgctc (500)
4	152923988–152924449	57	AmpliTaq Gold	tttgcctacgatatgaacaacc (50)	tttccaaatctgatccacctt (300)
5	152924334–152924567	58	AmpliTaq Gold	atgacaggctttgggggata (300)	gctgatccactgccttg (300)
6	152924404–152925092	58	AmpliTaq Gold	tggaaattctctgactagattg (300)	ataacgaacatggggactgg (50)
7	152924689–152924849	57	rTth XL	6FAM-ctgaggctgaggagaattgc (300)	tgagcctgcatgtcttcttt (300)
8	152925000–152925248	61	AmpliTaq Gold	ccattccatccaaatgttca (900)	gaggcgcacatgagggtgt (300)
9	152925036–152925735	59	AmpliTaq Gold	tgaaggatgtcaatggctgt (300)	gctgcagctgccttctagt (50)
10	152925598–152925823	61	AmpliTaq Gold	cttctcagaggggctcacc (300)	tttattcattcagtggtactattagg (900)
11	152925705–152926229	59	AmpliTaq Gold	catctccccaactagaaagg (300)	tggattttcattttgtgtgt (300)
12	152926138–152926801	61	AmpliTaq Gold	gtgtgcgtttacggacctct (500)	aggccagagggaccacaag (500)
13	152926644–152926893	59	AmpliTaq Gold	ccgcgtggagtcatttag (300)	aagaacgagggggtgagg (300)
14	152926753–152926935	61	AmpliTaq Gold:Pfu	cagagggcgcctgacattg (900)	gaagggtggatgaaagtgga (900)
15	152926844–152927233	60	AmpliTaq Gold:Pfu	<u>cgcccgccgcactgctccgctcctc</u> (900)	tccaggattacgctaagacc (300)
16	152926992–152927240	60	AmpliTaq Gold:Pfu	cgcccgccgccgccgccgccaccaggtgctgagtgactt (900)	<u>cgcccgccgccgcccttgc</u> caggattacg (900)
17	152927142–152927371	60	AmpliTaq Gold:Pfu	<u>cgcccgccgccgccgccgccaccatctccttctggagtg</u> (900)	<u>cgcccgccgccgccgccgctattt</u> ggaagcctttgag (900)
18	152927201–152927564	60	AmpliTaq Gold	tgcactgagctgggtcatta (50)	gcctccttgagatcacctg (50)
19	152927364–152927858	61	AmpliTaq Gold:Pfu	<u>cgcccgccgcccaaatagcatgggtgcctt</u> (900)	<u>cgcccgccgccgccctctgtg</u> cagtgctt (900)
20	152927634–152929316	68	TaKaRa	ctcttggaccggggaagagg (300)	agacttccagggccattc (300)
21	152927720–152928239	68	AmpliTaq Gold	gaaccacctcaggcatctca (600)	ggatgacccttgaggacattgc (600)
22	152928144–152928389	59	AmpliTaq Gold:Pfu	<u>cgggacgcctt</u> gaccactctaggecctt (300)	<u>cgggacgc</u> cagtgagctctgatggcaag (900)
23	152928225–152928837	62	AmpliTaq Gold	cctcaagggtcatccatttggga (500)	aaaattggattgtgatggccag (500)
24	152928461–152929131	64	AmpliTaq Gold	ctgcgactgtgccctgctaa (300)	tgagtgggtaaaaagtggcctgt (300)
25	152928950–152929272	60	AmpliTaq Gold	tgccctatctaagcatttcaca (50)	ccccaaactctggcacttcta (50)
26	152929098–152929674	60	AmpliTaq Gold:Pfu	<u>cgcccgccgccgccgccgctcc</u> cattgtacagacaggccact (900)	ctttcacggcttccatcag (900)
27	152929530–152929823	60	AmpliTaq Gold	gaatcgtggctctgagaacag (300)	catctcgtctgccctgtgt (300)
28	152929656–152929946	60	AmpliTaq Gold:Pfu	<u>cgcccgccgcctg</u> atggaagccgtgaaag (900)	<u>cgcccgccgc</u> acatgcagtgccacat (300)
29	152929886–152930198	64	AmpliTaq Gold	aactcagggcaaggagacg (600)	actgctcagagcccgaatcc (600)
30	152930034–152930326	61	AmpliTaq Gold	cccctacactggaggagga (300)	gactgtgggagccgctgt (50)
31	152930181–152930676	62	AmpliTaq Gold	<u>cgcccgccgccgccccattcggc</u> ctctgagcagtt (900)	ctgggtgctgctcatagc (300)
32	152968633–152969085	61	AmpliTaq Gold:Pfu	<u>cgcccgccgccgccgccgctcc</u> gccactgtgccagccactttt (50)	ctgggtgctgctcatagc (50)

¹Underlined residues were added to the primer sequence to improve PCR amplification.

Table 2. Polymorphic positions upstream of L/M opsin gene array

Nucleotide #	No. of subjects with		Nature of variation	N	
	Allele A	Allele B			
152923099	1	36	indel	37	
152923125	18	20	snp	38	
152923166	16	22	snp	38	
152923257	18	20	indel	38	
152923416	1	37	snp	38	
152923481	1	37	snp	38	
152923512	1	37	snp	38	
152923694	4	34	indel	38	
152923892	1	37	snp	38	
152924547	1	37	snp	38	
152924842	1	37	snp	38	
152924882	1	37	snp	38	
152924949	2	36	snp	38	
152925156	1	32	snp	33	
152925452	1	37	snp	38	
152925455	2	36	snp	38	
152926095	6	32	snp	38	
152926498	3	35	snp	38	
152926750	2	92	snp	94	
152927346	3	91	snp	94	
152928084	1	93	snp	94	
152928127	8	86	snp	94	
152928238	6	88	snp	94	
152928505	1	93	snp	94	
152928787	8	86	snp	94	
152928788	8	86	snp	94	
152929474	1	93	snp	94	
152929491	1	93	snp	94	
152929731	24	70	snp	94	
152929802	1	93	snp	94	
152929979	1	93	snp	94	
152930031	1	93	indel	94	
Upstream of M gene	152948808	1	94	snp	94
	152948869	1	94	snp	94
	152948922	1	94	indel	94

Indel = insertion or deletion; SNP = single nucleotide polymorphism.

pair region upstream of the L opsin coding sequence in our subject population (Table 2). Another three polymorphic positions, two SNPs and one insertion/deletion, occurred in the 452 base pairs upstream of the M-opsin coding sequence (Table 2). For each subject, we constructed a complete haplotype (a combination of polymorphisms) comprised of all of the polymorphic positions identified in this study, and we constructed two partial haplotypes. Haplotypes are given in Table 3. The complete haplotype for each subject was generated by stringing together the nucleotides present at each of the polymorphic positions listed in the first column of Table 2, and partial haplotypes were constructed by stringing together the nucleotides present at the polymorphic positions for two subregions. One subregion includes the known transcriptional control elements for the L and M opsin genes and extends from 152,926,644 to 152,923,076 (amplicons 13–31, Table 1) and from 152,968,633 to 152,969,085 (amplicon 32, Table 1), henceforth referred to as the proximal subregion. The second subregion spans a DNA segment that was previously reported to contain an SNP that correlated to the ratio of L:M opsin mRNA in donor retinas (Stamatoyannopoulos et al., 2005). This segment extends from

nucleotide 152,923,038 to 152,926,801, and will be referred to as the distal subregion. The rationale for splitting the haplotypes into two subregions for further analysis is two-fold. First, as is evident in Table 3, when evaluating all of the polymorphic positions in the complete haplotype, the number of different haplotypes is large with very low representation for each individual haplotype, and this is not amenable to statistical analysis. Second, the proximal subregion contains the known DNA regulatory elements that are involved in controlling transcription of the opsin genes and polymorphisms in this region are likely candidates for influencing the relative number of L versus M cones.

In the complete haplotypes (Table 3), the 32 polymorphic positions defined eight Caucasian and 11 African haplotypes. Africans and Caucasians shared the same most frequent haplotype, but its frequency was much lower in the African sample, only about 25%, compared to 50% of the Caucasian sample. No other overlap between African and Caucasian haplotypes occurred. Similar to other reports (i.e., Verrelli & Tishkoff, 2004), in comparison to Caucasians, there was more diversity among the complete African haplotypes, with nine of the 15 African subjects having a unique haplotype whereas only six of 18 Caucasian subjects had a unique haplotype (Table 3).

To evaluate the relationship between cone ratio and haplotype, we generated relatively high frequency haplotypes for each ethnic group by evaluating the proximal, and distal subregions separately. Any two individuals will differ on average at one in every 1,200 nucleotides, representing the accumulation of rare, random mutations (Venter et al., 2001). By reducing the size of the region and considering only the polymorphic sites that occur at relatively high frequency, we generated a data set that is amenable to statistical analysis. In these comparisons, nucleotide positions at which polymorphisms occurred in fewer than three people were excluded because the small numbers preclude statistical analysis (Table 3). For each ethnic group, this resulted in three proximal subregion haplotypes with at least five people per haplotype and three haplotypes with fewer than five people per haplotype. Further reduction of the proximal segment haplotypes excluding all but the six highest frequency polymorphisms in the proximal subregion for Africans and Caucasians combined allowed all but one Caucasian to be categorized into three proximal subregion haplotypes, and all Africans to be categorized into four haplotypes. These reduced proximal subregion haplotypes were extracted from the data in Table 3, by stringing together only the underlined nucleotides (the six highest frequency polymorphisms) for the proximal subregion. For each ethnic group, the frequency of each of the reduced haplotypes was calculated and a histogram of the data is shown in Fig. 1.

Inspection of Fig. 1A shows substantial differences in distributions for percentages of L cones for the different Caucasian haplotypes. The mean percent L cones for individuals with the GTCGGA haplotype is 64.6%, the mean for the GTCGGG haplotype is about 8% higher (72.6% L cones), and the GTTGGA haplotype is 6% higher than that (78.4% L cones). Thus, GTCGGA (low % L haplotype) versus GTTGGA (high % L haplotype) groups are a striking 13.8% different in percent L cones and this difference tests as significant with $p = 0.0194$ (Kruskal-Wallis nonparametric ANOVA) and $p < 0.05$ following the application of Dunn's correction for multiple comparisons. This difference of 13.8% L cones makes the difference between ratios of about 2 L for every M cone versus nearly 4 L for every M cone.

Fig. 1B shows that the distribution of the percent L cones among African haplotypes is more overlapping compared to Cau-

Table 3. Haplotypes¹

Caucasian Haplotypes	No.	Mean ² % L Cones	African Haplotypes	No.	Mean % L Cones
152923099–152948922 (Complete Region)			152923099–152948922 (Complete Region)		
()AG()CAT(TTTTC)AGCATGTTCTGGCTCGGGGGGGGCC()	9	71.1	()AG()CAT(TTTTC)AGCATGTTCTGGCTCGGGGGGGGCC()	4	57.6
()GACCAT(TTTTC)AGCATGTTCTGGCTCGGGGGAGGCC()	3	48.2	()GACCAT()AGCAGGTTACGGCTCGGGGGGGGCC()	2	
()AG()CAC(TTTTC)AGCATGTTCTGGCTCGGGGGGGGCC()	1		()AG()CAT(TTTTC)AGCATGTTCTGGCTCGGGGGAGGCC()	1	
()AG()CAT(TTTTC)AGCTTGTCTGGCTCGGGGGGGGCC()	1		()AGCCAT(TTTTC)AACATGTTATGGCACGTAGGGGGGCC()	1	
()AG()CAT(TTTTC)CGAATAACCTGGATCAGGGGGGGGCTC()	1		()GA()CAT(TTTTC)AGCATGTCATGGCACGTAGGGGGGCC()	1	
()GACCAT(TTTTC)AGCATGTTCTGGCTCGGGGGGGGCC()	1		()GACCAT()AGCATGTTATGGCACGTAGGGGGGCC()	1	
()GACCAT(TTTTC)AGCATGTTCTGGCTCGGGGGAGGCC()	1		()GACCAT()AGCATGTTCTGGCTCGGGGGAGGCC()	1	
()GGCCAT(TTTTC)AGCATGTTCTGGCTCGGGGGAGGCC()	1		()GACCAT(TTTTC)AGCATGTTACGGCACGTAGGGGGGCC()	1	
			()GACCAT(TTTTC)AGCATGTTCTGACTCGGGGGAGGCC()	1	
			()GACCAT(TTTTC)AGCATGTTCTGGCACGTAGGGGGGCC()	1	
			(GAGC)GGCGGT(TTTTC)AGCATGTTCTGGCTCGGGGGAGGCC()	1	
152926750–152930031 (Proximal Subregion)³			152926750–152930031 (Proximal Subregion)³		
<u>GGCTCGGGGGGGGCC()</u>	45	72.6	<u>GGCTCGGGGGGGGCC()</u>	9	59.5
<u>GGCTCGGGGGAGGCC()</u>	11	64.6	<u>GGCACGTAGGGGGGCC()</u>	8	61.7
<u>GGCTTGGGGAGGCC()</u>	5	78.4	<u>GGCTCGGGGGAGGCC()</u>	5	64.2
<u>AGCTCGGGGGGGGCC()</u>	2		<u>GACTCGGGGGAGGCC()</u>	2	
<u>GGATCAGGGGGGGCTC()</u>	1		<u>GACTCGGGGGAGGCC</u>	1	
<u>GGCTCGGGGAGGCC()</u>	1		<u>GGCTCGGGGGAGGCC()</u>	1	
<u>GGCTCGGGGGGGCCT()</u>	1		<u>GGCTCGGGGGGGC()</u>	1	
<u>GGCTTGGGAGGGGCC()</u>	1				
152923099–152926498 (Distal Subregion)			152923099–152926498 (Distal Subregion)		
()AG()CAT(TTTTC)AGCATGTTCT	9	71.1	()AG()CAT(TTTTC)AGCATGTTCT	5	55.6
()GACCAT(TTTTC)AGCATGTTCT	5	62.2	()GACCAT()AGCAGGTTAC	2	
()AG()CAC(TTTTC)AGCATGTTCT	1		()GACCAT(TTTTC)AGCATGTTCT	2	
()AG()CAT(TTTTC)AGCTTGTCT	1		()AGCCAT(TTTTC)AACATGTTAT	1	
()AG()CAT(TTTTC)CGAATAACCT	1		()GA()CAT(TTTTC)AGCATGTCAT	1	
()GGCCAT(TTTTC)AGCATGTTCT	1		()GACCAT()AGCATGTTAT	1	
			()GACCAT()AGCATGTTCT	1	
			()GACCAT(TTTTC)AGCATGTTAC	1	
			(GAGC)GGCGGT(TTTTC)AGCATGTTCT	1	

¹The nucleotide present at each position indicated in the first column of Table 2 are strung together to give a haplotype for each Caucasian and African subject. Within the haplotypes, parentheses () indicate positions at which the variation was an insertion or a deletion (indel, Table 2), with empty parentheses denoting a deletion and the contents within parentheses listing the nucleotides inserted. The column labeled No. indicates the number of individuals in the sample that had the indicated haplotype. The complete region gives all positions listed in the first column of Table 2 and was obtained only for a subset of subjects (18 Caucasians and 15 Africans), the proximal subregion includes the positions in Table 2 from 152926750 through 152930031, and the distal subregion includes the haplotypes for positions 152923099 through 152926498. Proximal subregion haplotypes were obtained from 67 Caucasians and 27 Africans and distal subregion haplotypes were obtained from 18 Caucasians and 15 Africans.

²Mean %L cones is given as the average of the %L cones estimated for the individuals of each ethnic group the given haplotype. For each haplotype the mean %L cones equals $100 * (\Sigma \%L \text{ cones}) / \text{number individuals with the haplotype}$.

³Haplotypes in Figure 1 are constructed from the underlined positions in the proximal region haplotypes which are 152,927,346; 152,928,127; 152,928,787; 152,928,788; 152,929,731.

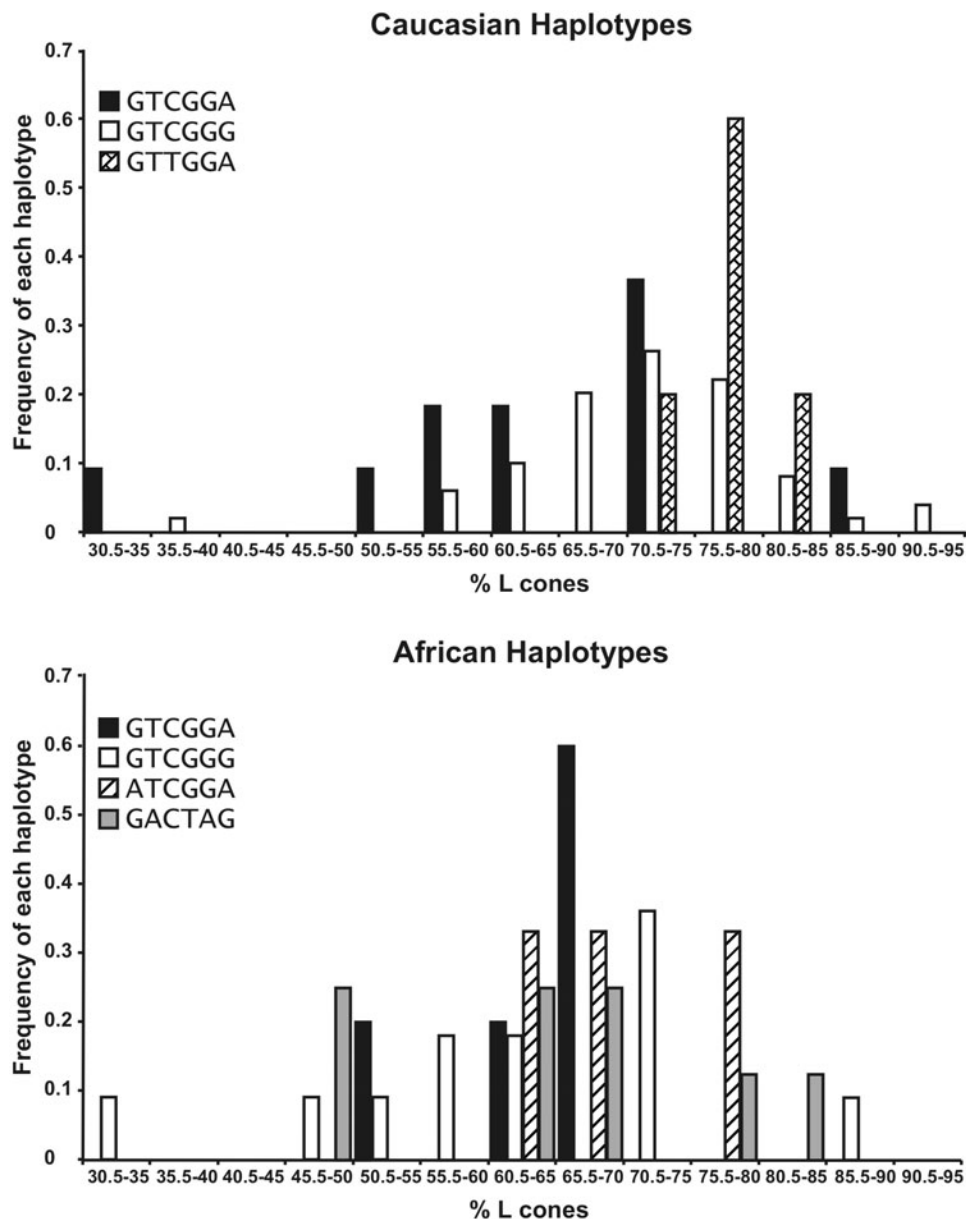


Fig. 1. Histograms of the percent L cones ($100 * L / (L + M)$) for the most common reduced haplotypes for the proximal subregion of the X-chromosome opsin gene locus for Caucasians (A) and Africans (B). The reduced haplotypes include X-chromosome nucleotide positions 152,927,346; 152,928,127; 152,928,787; 152,928,788; 152,929,731. Haplotype frequencies are reported on the Y-axis as the number of people with the range of percent L cones given on the X-axis divided by the total number of people with the same haplotype.

casians. Consistent with this, the differences among African haplotypes do not test as being statistically different ($p = 0.6607$, Kruskal-Wallis nonparametric ANOVA). However, as we have reported earlier (McMahon et al., 2008), there are striking differences in cone ratio between Africans and Caucasians. Part of the difference can now be explained by the fact that the Caucasian “high % L haplotype” GTTGGA is absent from the African sample, while the Caucasian “low % L haplotype” GTCGGA is highly represented among the Africans; yet, these two differences do not completely explain the differences in cone ratio between the Africans and Caucasians because, in our sample, Africans versus Caucasians who share the most common haplotype in the proximal subregion (GGCTCGGGGGGGGCC), (Table 3) still differ sig-

nificantly in mean percentage of L cones (unpaired t test, $p = 0.0011$; 72.6% and 59.5% L cones for Caucasian and African, respectively). This suggests that additional nucleotide differences, such as those that occur in the distal subregion, also tune the relative numbers of L and M cones. Incidentally, one other proximal subregion haplotype is shared by Caucasians and Africans, GGCTCGGGGGGAGGCC, but the median percentage of L cones did not differ between these two relatively small groups ($p = 0.7427$; 64.6% versus 64.2% L cones).

The distal subregion spans a segment of DNA that was reported previously to contain a polymorphism that was significantly associated with the ratio of L:M cones as estimated from the ratio of L versus M opsin messenger RNA isolated from donor retinas (Stama-

toyannopoulos et al., 2005). The polymorphism was reported to lie 7.4 kb upstream of the L opsin gene and to occur at high frequency in the population. We found two polymorphic nucleotide positions, 152,923,125 and 152,923,166, in the distal subregion that lie 7.4 kb upstream of the L opsin gene, and that occur at high frequency (Table 2); however, the difference in the mean percentage of L cones for the subjects that differ at these two positions was not statistically significant in our sample (Mann Whitney U test $p = 0.2786$).

Discussion

How might the differences among haplotypes affect the percentage of L cones in the retina? Polymorphisms in the regions examined in this study have a high potential for altering the binding affinity for proteins involved in regulating transcription, and thereby in the choice made in each photoreceptor of which X-chromosome opsin gene to express. Differences among the haplotypes may tip the balance more or less in favor of expressing L opsin, and thus are likely candidates for producing subtle differences such as the difference between the intermediate versus the high and low % L cones phenotypes. Recently, evidence has accumulated indicating that the difference between L versus M cones is decided simply by which of the X-chromosome cone opsin genes is chosen for expression (Smallwood et al., 2002; Knoblauch et al., 2006). DNA polymorphisms in the region of the opsin genes containing elements involved in controlling gene expression and silencing can influence the transcriptional output of each of the X-chromosome opsin genes. Since each polymorphism can either act neutrally, synergistically, or counter to each other polymorphism, the contribution of individual polymorphisms can be difficult to detect and to quantify, especially if the contribution is relatively minor. Difficulties in detection are compounded by the relatively low frequency of each haplotype.

In this study, after correcting for multiple comparisons, one polymorphism was detected as significantly correlated to the percentage of L cones. For the Caucasian proximal region haplotypes, pair wise comparisons showed a significant difference in the mean percentage of L cones between GTCGGA (relatively low % L cones) versus GTTGGA (relatively high % L cones). The nucleotide position that differs between these haplotypes is C versus T at 152,928,238. This serves as an example of an association between a nucleotide difference and cone ratio that is predicted by the hypothesis that nucleotide polymorphisms upstream of the X-chromosome opsin gene array tune L:M cone ratio.

In summary, this study provides insight into the complex nature of the biological control of the relative number of L versus M cones. The one statistically significant difference identified here supports the principle that the nucleotide differences in the region of the opsin genes are responsible for determining cone ratio. Because of the high degree of variability, identifying which other nucleotides are involved will require testing many more subjects. The present analysis did identify one other candidate nucleotide. High and low % L cones haplotypes differed from the intermediate % L cones haplotype in an A versus G nucleotide at position 152,929,731. After correcting for multiple comparisons in our data set, the difference in the mean percent L cones was not quite statistically significant between the GTCGGG haplotype associated with the intermediate % L cones and either the high (GTTGGA) or low (GTCGGA) % L cones haplotypes; however,

this nucleotide merits further study in future efforts to identify genetic changes associated with variability in cone ratio.

Acknowledgments

This work was supported by National Eye Institute Grants F32EY014789 (KLG), T32EY014537, and R01EY09303, P30EY01961 and by an unrestricted Research to Prevent Blindness (RPB) grant to the department of ophthalmology. Jay Neitz was the recipient of an RPB Senior Scientific Investigator Award.

References

- BUMSTED, K., JASONI, C., SZEL, A. & HENDRICKSON, A. (1997). Spatial and temporal expression of cone opsins during monkey retinal development. *Journal of Comparative Neurology* **378**, 117–134.
- CARROLL, J., NEITZ, M. & NEITZ, J. (2002). Estimates of L:M cone ratio from ERG flicker photometry and genetics. *Journal of Vision* **2**, 531–542.
- DARTNALL, H.J.A., BOWMAKER, J.K. & MOLLON, J.D. (1983). Human visual pigments: Microspectrophotometric results from the eye of seven persons. *Proceedings of the Royal Society of London B* **220**, 115–130.
- DEVRIES, H.L. (1946). Luminosity curve of trichromats. *Nature (London)* **157**, 736–737.
- FISHER, R.A. (1918). The correlation between relatives under the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399–433.
- GUNTHER, K.L., NEITZ, J. & NEITZ, M. (2006). A novel mutation in the short-wavelength sensitive cone pigment gene associated with a tritan color vision defect. *Visual Neuroscience* **23**, 403–409.
- HAGSTROM, S.A., NEITZ, M. & NEITZ, J. (2000). Cone pigment gene expression in individual photoreceptors and the chromatic topography of the retina. *Journal of the Optical Society of America A-Optics Image Science and Vision* **17**, 527–537.
- HOFER, H., CARROLL, J., NEITZ, J., NEITZ, M. & WILLIAMS, D.R. (2005). Organization of the human trichromatic cone mosaic. *Journal of Neuroscience* **25**, 9669–9679.
- KNOBLAUCH, K., NEITZ, M. & NEITZ, J. (2006). An urn model of the development of L/M cone ratios in human and macaque retina. *Visual Neuroscience* **23**, 591–596.
- MCMAHON, C., CARROLL, J., AWUA, S., NEITZ, J. & NEITZ, M. (2008). The L:M cone ratio in males of African descent with normal color vision. *Journal of Vision* **8**, 1–9.
- POKORNY, J., SMITH, V.C. & WESNER, M.F. (1991). Variability in cone populations and implications. In *From Pigments to Perception: Advances in Understanding Visual Processes*, ed. Valberg, A. & Lee, B.B., pp. 23–33. New York: Plenum Press.
- ROORDA, A. & WILLIAMS, D.R. (1999). The arrangement of the three cone classes in the living human eye. *Nature* **397**, 520–522.
- RUSHTON, W.A.H. & BAKER, H.D. (1964). Red/green sensitivity in normal vision. *Vision Research* **4**, 75–85.
- SMALLWOOD, P.M., WANG, Y. & NATHANS, J. (2002). Role of a locus control region in the mutually exclusive expression of human red and green cone pigment genes. *Proceedings of the National Academy of Sciences: USA* **99**, 1008–1011.
- STAMATOYANNOPOULOS, J., THURMAN, R., NOBLE, W., KUTYAVIN, T., SHAFER, A., DORSCHNER, M. & DEEB, S.S. (2005). Quantitative phenotype variation in long- and middle-wave cone photoreceptors due to polymorphism in an upstream insulator element. In *American Society of Human Genetics*, Abstract #273. American Society of Human Genetics, Salt Lake City, Utah.
- STRACHAN, T. & READ, A.P. (2000). *Human Molecular Genetics 2*. New York: John Wiley & Sons, Inc.
- VENTER, C.J., ADAMS, M.D., MYERS, E.W., LI, P.W., MURAL, R.J., SUTTON, G.G., AMANATIDES, P., BALLEW, R.M., HUSON, D.H., WORTMAN, J.R., ZHANG, Q. & KODIRA, C.D. (2001). The sequence of the human genome. *Science* **291**, 1304–1351.
- VERRELLI, B.C. & TISHKOFF, S.A. (2004). Signatures of selection and gene conversion associated with human color vision variation. *American Journal of Human Genetics* **75**, 363–375.