Genetic Basis of Differential Opsin Gene Expression in Cichlid Fishes

Karen L Carleton, **Christopher M Hofmann**, **Christie Klisz**, **Zil Patel**, **Lauren M Chircus**, **Lauren H Simenauer**, **Natasha Soodoo**, **R Craig Albertson**[*], and **Jennifer R Ser**
Department of Biology, University of Maryland, College Park MD 20742

Abstract

Visual sensitivity can be tuned by differential expression of opsin genes. Among African cichlid fishes, seven cone opsin genes are expressed in different combinations to produce diverse visual sensitivities. To determine the genetic architecture controlling these adaptive differences, we analyzed genetic crosses between species expressing different complements of opsin genes. Quantitative genetic analyses suggest that expression is controlled by only a few loci with correlations among some genes. Genetic mapping identifies clear evidence of trans-acting factors in two chromosomal regions that contribute to differences in opsin expression as well as one cis-regulatory region. Therefore, both cis and trans regulation are important. The simple genetic architecture suggested by these results may explain why opsin gene expression is evolutionarily labile, and why similar patterns of expression have evolved repeatedly in different lineages.

Keywords

gene expression; opsin; cis regulation; trans-acting factor

Introduction

There has been considerable debate regarding the molecular mechanisms underlying organismal diversity. King and Wilson (1975) first proposed that changes in gene regulation were critical to phenotypic differences among closely related species. More recent work has suggested that changes in cis regulatory modules are the most common basis for evolutionary changes in development (Carroll, 2008, Wray, 2007). Others maintain that structural changes in proteins are the most important mechanism of evolutionary change (Hoekstra & Coyne, 2007).

With modern molecular tools, we can study the genetic architecture underlying a broad range of phenotypes. While many of the previous studies have focused on morphology or pigmentation, adaptive variation exists in many other traits. Variation in sensory systems is critical for foraging, mating, and predator avoidance (Endler, 1992). Sensory systems are optimized for particular environments, and organisms that inhabit different environments may diverge in sensitivity. Sensory divergence can lead to speciation through the process of sensory drive (Boughman, 2002, Ryan & Rand, 1993). Recently, we have shown that both coding sequence changes and differences in gene expression can alter visual sensitivities (Hofmann et al., 2009). However, we do not know if changes in gene regulation are the result of cis- or trans-regulatory elements. Understanding the genetic basis of sensory

Corresponding author: Karen Carleton, Department of Biology, 1210 Biology / Psychology Bldg #144, University of Maryland, College Park, MD 20742, phone: 301-405-6929, fax: 301-314-9358, kcarleto@umd.edu.
[*]Department of Biology, Syracuse University, Syracuse NY

systems is important to understanding how organismal diversity arises and how rapidly it can change. Further, such diversity is important as it can lead to adaptation as well as help drive speciation.

African cichlid fishes are a model for organismal diversity and rapid speciation (Fryer & Iles, 1972, Kocher, 2004, Seehausen, 2006, Salzburger, 2009). While they are recognized for the diversity of their trophic morphology (Liem, 1991, Albertson et al., 2003a), dentition (Fraser et al., 2008, Fraser et al., 2009) and color patterns (Konings, 2007, Seehausen, 1996), we have found that they also show incredible diversity in visual sensitivity, which is determined by the visual pigments in their photoreceptors (Hofmann et al., 2009, Carleton & Kocher, 2001, Parry et al., 2005, Jordan et al., 2006).

Visual pigments are composed of an opsin protein bound to a chromophore such as 11-cis retinal (Wald, 1935). Cichlids have seven cone opsin genes which produce visual pigments with sensitivities across the spectrum from ultraviolet to red wavelengths. These cone opsins include short wavelength sensitive (SWS), rhodopsin like (RH2), and long wavelength sensitive (LWS) genes. The peak wavelengths (λ_{max}) of the expressed proteins in the riverine cichlid *Oreochromis niloticus* are: SWS1 (360 nm), SWS2B (423), SWS2A (456), RH2B (472 nm), RH2Aβ (518 nm), RH2Aα (528 nm), and LWS (561 nm) (Spady et al., 2006). Because the RH2Aα and RH2Aβ genes are genetically and functionally similar, we group them together into the RH2A class, so that the seven opsin genes are organized into six distinct classes.

In previous work, we have found that cichlids utilize two different mechanisms for tuning their visual pigments. First, the cichlid opsin coding sequences vary to produce small spectral shifts in the corresponding visual pigments. These shifts occur primarily in the genes at the very shortest and longest wavelengths, SWS1 and LWS (Hofmann et al., 2009). Shifts of 5-10 nm in the LWS gene help to drive speciation between red and blue cichlid species in Lake Victoria through sensory drive (Seehausen et al., 2008, Terai et al., 2006, Carleton et al., 2005). Second, cichlid opsin genes are differentially expressed to create large spectral shifts and a diversity of visual palettes. Gene expression differs considerably across the cichlid phylogeny within Lake Malawi with differences between genera and even between species within a given genus (Hofmann et al., 2009). This evolutionary lability suggests that it is genetically easy to alter opsin gene expression to spectrally tune visual sensitivity to species ecology. Since opsin sequence tuning seems significant for only SWS1 and LWS, where shifts are 5-10 nm, the majority of spectral shifts occur through changes in opsin gene expression. In Lake Malawi, we have found three gene combinations to be most common. We name these gene sets by the spectral sensitivity of the shortest wavelength gene: UV (SWS1, RH2B, RH2A), violet (SWS2B, RH2B, RH2A) and blue (SWS2A, RH2A, LWS).

The expression of cichlid opsins occurs in two morphologically distinct cones: single and double. These are highly ordered in a retinal mosaic which in cichlids is a central single cone surrounded by four pairs of double cones (Fernald, 1981). Typically the central cone expresses a short wavelength sensitive opsin (either SWS1, SWS2B or SWS2A). The surrounding double cones contain the longer wavelength pigments from expression of RH2B, RH2A or LWS (Parry et al., 2005, Carleton et al., 2008, Jordan et al., 2006, van der Meer & Bowmaker, 1995). Differential opsin gene expression therefore produces a patterned retina, which varies only in the combination of pigments. This patterning might be controlled by cis regulatory elements, in a manner similar to the regulation of color patterns found in *Drosophila* wings (Gompel et al., 2005, Prud'homme et al., 2006).

Studies in other fish species have identified both cis and trans regulatory factors that affect opsin expression. SWS1 opsin expression in salmon responds to retinoic acid and thyroid hormone (Browman & Hawryshyn, 1994a, Browman & Hawryshyn, 1992, Browman & Hawryshyn, 1994b). This is consistent with studies that suggest that thyroid hormone is important in the differentiation of mouse cone photoreceptors (Kelley et al., 1995, Ng et al., 2001, Applebury et al., 2007). The expression of these hormones normally varies during development in fish and amphibians and so act as trans-regulator factors for opsins as well as other genes (Jones et al., 2003, Paris & Laudet, 2008, Marchand et al., 2004). Analysis of cis regulatory regions have identified several potential transcription factor binding sites in zebrafish (Luo et al., 2004, Tsujimura et al., 2007) and salmon (Dann et al., 2004). However, on the whole it has been difficult to use genetic approaches because model systems such as zebrafish (Schmitt et al., 1999) and mouse (Neitz & Neitz, 2001) do not show sufficient variation in opsin expression.

To unravel the molecular mechanism by which opsin gene expression changes, we have made two genetic crosses between cichlid species which differ in the expressed opsin gene combinations. We quantify variation in the F_2 generation to estimate how many genes control opsin expression. We also test for association between genetic markers located both cis and trans to the opsin genes. This study provides preliminary data for a larger QTL study and so utilizes a smaller set of individuals. As such, our goal is not to identify all the controlling factors. Rather, here we parameterize the phenotype to establish whether there are few or many controlling factors in order to assess if it is possible to map these factors. Further, we test whether such factors are cis to the opsin genes as has been found in many other phenotypes. Finding partial evidence for cis regulation, we perform a preliminary genome scan and find evidence that trans acting factors are key to opsin gene expression.

Materials and methods

1. Genetic architecture of factors controlling opsin gene expression

1.1 F2 crosses—Two intergeneric crosses were made between cichlid species with different opsin expression. The first cross was between *Dimidiochromis compressiceps* and *Copadichromis eucinostomus* (DC cross). These species use the blue (SWS2A, RH2A, LWS) and the UV (SWS1, RH2B, RH2A) gene sets, respectively. This cross between a predator and a zooplanktivore was difficult to make, and we were only able to rear one F_2 family of 33 individuals to 15 days post fertilization. The second cross was between *Tramitichromis intermedius* and *Aulonocara baenschi* (TA cross). These species use the blue (SWS2A, RH2A, LWS) and the violet (SWS2B, RH2B, RH2A) gene sets, respectively. This cross was prolific, hardy, and easy to raise. We crossed two male *T. intermedius* and five full sib female *A. baenschi* to create five F_1 families. The F_1 families were then intercrossed to raise numerous F_2 families. For this study, we analyzed 49 F_2 individuals which came from 5 F_2 families derived from 2 F_1 families. All individuals were raised to at least six months in age (sexual maturity).

1.2. Phenotyping—Individuals were euthanized with MS-222, the eyes removed and the dissected retina placed in RNAlater. Total RNA was extracted (Qiagen RNeasy kit) and reverse transcribed (Superscript III, Invitrogen). This cDNA template was then used in quantitative RT-PCR to measure the relative expression of all six opsin classes (SWS1, SWS2B, SWS2A, RH2A, RH2B and LWS) according to our previous methods (Spady et al., 2006, Carleton et al., 2008). The expression of these six gene classes is normalized to give expression of each class relative to the total:

$$f_i = \frac{T_i}{T_{all}} = \frac{\frac{1}{(1+E_i)^{C_{ti}}}}{\sum \frac{1}{(1+E_i)^{C_{ti}}}}$$

(1)

where T_i is the relative amount of each gene i, E_i is the PCR efficiency of that gene, and C_{ti} is the gene's critical cycle number. Here, f_i is the fraction that each gene contributes to total opsin expression.

Gene expression can also be normalized within the single and double cones since single cones express only SWS1, SWS2B and SWS2A, while double cones express RH2B, RH2A and LWS. So for example, the fraction of SWS1 expression in single cones is given by:

$$f_{SWS1,SC} = \frac{T_{SWS1}}{T_{SWS1} + T_{SWS2b} + T_{SWS2a}}$$

(2a)

while the fraction of LWS expression in double cones is given by

$$f_{LWS,DC} = \frac{T_{LWS}}{T_{RH2b} + T_{RH2a} + T_{LWS}}$$

(2b)

These normalized measures were used to make triangle plots of single and double cone gene expression.

In order to visualize the expression data, we estimated the relative sensitivity of the single and double cones. This utilized known peak sensitivities of each gene weighted by the relative gene expression in each individual (Carleton et al., 2008):

$$\lambda_{SC} = \frac{f_{SWS1}\lambda_{SWS1} + f_{SWS2b}\lambda_{SWS2b} + f_{SWS2a}\lambda_{SWS2a}}{f_{SWS1} + f_{SWS2b} + f_{SWS2a}}$$

(3a)

$$\lambda_{DC} = \frac{f_{RH2b}\lambda_{RH2b} + f_{RH2a}\lambda_{RH2a} + f_{LWS}\lambda_{LWS}}{f_{RH2b} + f_{RH2a} + f_{LWS}}$$

(3b)

This is not meant to imply that there are actual single and double cones with a sensitivity equal to these values. Rather, it is a convenient way to convert gene expression into biologically relevant wavelengths which can then be visualized in a two dimensional space. Gene specific peak wavelengths are taken from expressed opsin proteins for *Metriaclima zebra* (λ_{SWS1}=368 nm, λ_{SWS2B}=423 nm, λ_{RH2B}=484 nm, $\lambda_{RH2A\alpha}$=528 nm (Parry et al., 2005)) and for *Oreochromis niloticus* (λ_{SWS2A}=455 nm, λ_{LWS}=561 nm (Spady et al., 2006)). These values are very similar to the average obtained from MSP data for 19 species from Lake Malawi (λ_{SWS1}=371±8 nm, λ_{SWS2B}=418±5 nm, λ_{SWS2A}=452±3 nm, λ_{RH2B}=482±5 nm, λ_{RH2A}=527±7 nm, and λ_{LWS}=565±5 nm) (reviewed in Carleton, 2009). Although we group the RH2Aα and β together, these have similar protein expression peak wavelengths in both *M. zebra* and *O. niloticus* ($\lambda_{RH2A\alpha}$=528 nm, $\lambda_{RH2A\beta}$=518 nm).

Microspectrophotometry data for the Malawi species show green cones with an average λ_{max}=527 nm, which suggests that the RH2Aα gene predominates.

1.3. Number of genes controlling expression—The Castle-Wright estimator was used to calculate a lower bound to the number of factors controlling expression of each opsin gene (Lande, 1981, Lynch & Walsh, 1998, Wright, 1968). This method makes several assumptions, including that one parent's alleles increase the trait and the other's alleles decrease the trait, the alleles act additively, loci are unlinked, and all loci have equal allelic effects. These assumptions may not hold true in outbred crosses, such as those used here, and so this method provides only a lower limit to the number of controlling factors (Zeng, 1992, Zeng et al., 1990). We applied the Castle-Wright equation as formalized by Lande (1981):

$$n_e = \frac{(\mu_A - \mu_B)^2 - \sigma_A^2 - \sigma_B^2}{8(\sigma_{F2}^2 - \sigma_{F1}^2)}$$

(4)

where n_e is the number of factors controlling each gene, μ is mean gene expression in parental species A or B, and σ is the variance in expression of the parentals (A, B), F_1 or F_2. We estimated numbers from the fraction of the single or double cone expression ($f_{i, SC}$ and $f_{i, DC}$). For the DC cross, we phenotyped two individuals of each parental species as well as 33 F_2 which were 15 dpf. We did not have any F_1 to phenotype and so have used the larger of the parental variances for each gene as an estimate for σ_{F1}. This could lead to further underestimates in gene numbers, beyond the caveats already inherent in the estimator (Lynch & Walsh, 1998). For the TA cross, we phenotyped two *T. intermedius* and ten *A. hansbaenschi* individuals, 28 F_1 and 48 F_2. All were phenotyped at greater than six months of age, when they are considered adult.

We were able to estimate additional parameters for the TA cross where we have a good estimate of F_1 variance. First, we estimated the variance of n_e from (Lynch and Walsh, 1998 p. 234):

$$\text{var}(n_e) = n_e^2 \left[\frac{4(\sigma_A^2 + \sigma_B^2)}{(\mu_A - \mu_B)^2} + \frac{\text{var}(\text{var } S)}{\text{var}(S)^2} \right]$$

(5)

where

$$\text{var}(S) = \sigma_{F2}^2 - \sigma_{F1}^2$$
$$\text{var}(\text{var}(S)) = \frac{2(\sigma_{F2}^2)^2}{n_{F2} + 2} + \frac{2(\sigma_{F1}^2)^2}{n_{F1} + 2}$$

We were also able to estimate the degree of dominance in the TA cross from the means of the F_1 and the parentals. Here dominance is calculated as $d = (\mu_{F1} - \mu_A)/(\mu_T - \mu_A) - 0.5)$, where μ_{F1} is the mean of the F_1 and μ_T and μ_A are the parental means. This was done separately for each opsin gene. The Castle-Wright equation assumes additivity of allelic effect. Dominance effects can be included and act to increase the effective number of factors (Wright, 1968, p. 384). If $h = d + 0.5$, the number of factors with dominance are:

$$n_{e,dom}=[\,1.5-(2h(1-h)]n_e$$

This effect was taken into account in estimates of n_e for the TA cross.

2. Genetic mapping of factors controlling opsin gene expression

2.1 Test of cis regulation in both DC and TA crosses—In order to determine whether opsin gene expression is controlled by cis regulatory factors, we genotyped microsatellite markers near the opsin genes. The opsin genes are located in three genomic regions (Lee et al., 2005). SWS1 is located on linkage group (LG) 17. RH2B – RH2Aα – RH2Aβ are in a 26 kb tandem array on LG5. SWS2A – SWS2B – LWS is an 18kb tandem array also on LG5, though approximately 30 cM away from the RH2 array. The location and primers used to amplify each marker are given in Table 1. The UV marker used to genotype the SWS1 gene is in an intron of that gene and the same marker worked for both crosses. The other two opsin clusters (RH2 and SWS2-LWS) required different primer sets in each cross to successfully amplify a closely linked marker in an intergenic region. The most distant marker (redblue4) was 22 kb from the end of the LWS gene and therefore 38 kb from the beginning of the SWS2A – SWS2B – LWS array. These distances are still quite close genetically. The cichlid genetic map is 1300 cM for 10^9 bp such that 1 cM= 750 kb (Lee et al., 2005). In scoring 48 individuals (96 meioses), we expect fewer than 0.05 recombinants between the redblue4 marker and the beginning of SWS2A. These markers are therefore in tight linkage with the entire region.

2.2 Search for trans-acting regions—We performed a low pass search in the DC cross to identify chromosome regions which might control opsin expression. We used a combination of randomly selected markers across the genome, as well as a comparative approach to select fish chromosomes identified as containing candidate genes thought to be important for opsin expression (thyroid hormone receptors, retinoic acid receptors). Markers were selected either from the *Oreochromis niloticus* map (Lee et al., 2005) or a map made for two Malawi rock-dwelling cichlids (Albertson et al., 2003a). This included thirteen markers, three that were cis to the opsin genes and ten that were located across the genome, in trans to the opsin regions. Additional markers were tested in regions showing positive association, including 3 on LG4 and 11 on LG13. The opsin markers as well as markers in the regions of association were then tested in the TA cross.

2.3 Single marker association—Association between each marker and opsin expression was tested using analysis of variance (ANOVA). Individuals were grouped by genotype and then an ANOVA was performed to test whether gene expression for each gene differed by genotype. For the DC cross, this was done for expression of all six genes. In the TA cross, we performed single marker association for expression of all five opsin genes, leaving out SWS1, which was not expressed. Since each gene is a unique data set, those tests are independent. However, each gene is tested for association at multiple markers and so we need to correct for multiple tests. We make the additional assumption that linked markers are not independent since their genotypes are tightly correlated. The linked markers on LG4 (UNH2126, UNH2156) and LG13 (GM507, UNH871, GM641) were therefore treated as single tests. For each gene, we use sequential Bonferroni correction to test for significance (Hochberg, 1988). The marker tests are ordered by significance value and compared to the significance threshold determined by:

$$P_j=\alpha/(m-j+1)$$

where α=0.05 and j varies from 1 up to m, the number of markers/regions tested.

2.4 Map order—For the markers which were associated with opsin expression and located on LG13, we determined the map order for both the DC and the TA cross and compared that with the tilapia map order of Lee et al (2005). The DC / TA map order was determined from the recombination rates between markers along with the map order giving the largest likelihood.

The likelihood or LOD score for linkage between two markers is given by:

$$LOD = \frac{(1-\theta)^{NR}\theta^R}{0.5^{R+NR}}$$

where R is the number of recombinants between the two markers, NR is the number of nonrecombinants and θ is the recombination fraction = R / (R+NR). The LOD score for a given map order of three markers is just the sum of the LOD score for the two intervals, if interference is neglected (Ott, 1991). For the LG13 region, we are interested in the map order of GM507, UNH871, and GM641. Because there are no differences in the degrees of freedom between marker orders, we can not use a likelihood ratio test to assess order. However, we can calculate the probability of marker order 2 versus marker order 1 from

$$\Pr(\text{order2 vs order1}) = \exp(LOD_2 - LOD_1)$$

where LOD_1 and LOD_2 are taken as the scores for the tilapia map order and the Malawi cross (TA or DC) map order, respectively (Liu, 1998).

Results

3.1 a Gene expression in DC cross

Opsin gene expression was highly variable among the 33 F_2 of the DC cross, as shown in the triangle plots (Figure 1). F_2 expression for RH2B and RH2A varied between the extreme values of the parentals. F_2 expression for SWS1 and SWS2A fell short of the parental range. This was most significant for SWS2A, where F_2 expression was very low (max of 6.8%), in spite of strong expression in the *D. compressiceps* parent (20.15%). In contrast, expression of SWS2B was transgressive. The F_2 showed more expression (0.6 to 28.9%) than either of the parental species (0 and 2.4%). This is seen in the triangle plot of Fig 1A, where the parentals fall close to either the SWS1 or SWS2A vertices and yet the F_2 vary along the SWS1 – SWS2B axis.

3.1b Gene expression in the TA cross

Gene expression was also highly variable in the TA cross. The expression of the SWS2A and SWS2B in single cones varied across the full range defined by the parentals. As expected from the lack of SWS1 in the parentals, there was no significant SWS1 expression in either the F_1 or F_2. Expression for the double cone genes (RH2B, RH2A and LWS) extended somewhat beyond that of the parentals as shown in the triangle plots of Figure 1B. However, we do not see the transgressive expression in this cross that we observed in the DC cross.

3.1c Gene expression in wavelength space

Plots of gene expression were made for each cross in single / double cone wavelength sensitivity space. This space allows us to look at correlations between the single and double cones and to compare the crosses to our previous work (Hofmann et al., 2009, Carleton et al., 2008). As shown in Figure 2, the F_2 of both crosses varied across the range set by the values of the parental species. This is true for both single and double cone wavelengths. This large variation in the F_2 is consistent with the idea that a small number of genes control opsin expression.

3.2 Correlations among genes

Expression of each opsin was normalized to total opsin expression. This normalization will cause some inherent gene correlations, particularly between the highly expressed double cone genes, RH2B, RH2A and LWS. Here we note correlations between genes in the single and double cones, which are unlikely to result from normalization (Table 2). For the DC cross, there is a positive correlation between expression of SWS1 and RH2B and a negative correlation of SWS1 with SWS2B and LWS. For the TA cross, RH2B is positively correlated with SWS2B and negatively correlation with SWS2A. The single cone – double cone sensitivity plot (Figure 2) also shows evidence of gene correlations. Individuals with short wavelength single cones tend to have short wavelength double cones, while those with long wavelength single cones have long wavelength double cones. This is most obvious for the DC cross, where the regression coefficient for single and double cone wavelength is 0.55 ($F_{1,31}$=38.2; p=7.3 × 10^{-7}). However, the regression coefficient for the TA cross is only 0.06 ($F_{1,47}$=3.12 p=0.084).

3.3 Castle Wright estimates of genetic factors

Based on the gene expression differences, the parentals of both the TA and DC cross differ in gene expression by substantial amounts. For the TA cross, the parental differences given in standard deviations is: SWS2b 37.8, SWS2a 152, RH2b 7.6, RH2A 16.3, and LWS 41.1. These large differences support that the alleles of the two parentals work in opposite directions, and favor the applicability of the Castle-Wright estimator. However, the other caveats of additivity, linkage, and differential parental contributions can still confound these results (Zeng, 1992, Zeng et al., 1990).

We have estimated the number of factors controlling opsin expression using the parental means and the variances of each generation (Table 3). For the DC cross, we use the larger of the F_0 variances as an estimate of variances in F_1. For the majority of the genes in both crosses, we estimate that 1-2 genetic factors control the differences in expression between species. This estimate is a lower bound, but it supports the idea that a small number of genes control opsin gene expression.

There are two exceptions which suggest more complex genetic control. The first is the SWS2A gene in the DC cross where we estimate over 150 genes control expression. This is likely an artifact because the F_2 do not show significant expression of this gene, leading to a skewed estimate of the number of genes. The second is the LWS gene in the TA cross where we estimate four genes control expression.

3.4 Dominance

There was some evidence of dominance in the control of gene expression. This was estimated in the TA cross where we have expression data for the F_1. Values were d= 0.15 (SWS2B), 0.17 (SWS2A), -0.29 (RH2B), -0.35 (RH2A) and -0.33 (LWS). Here, the single cone genes have positive values, and are closer to *Aulonocara* gene expression levels, while the double cones have negative values, and are shifted towards the levels of gene expression

found in *Tramitichromis*. This suggests a small but significant effect of dominance in the expression of these genes. Including dominance in the estimate of controlling factors only slightly increases their values, and we still estimate 1-2 factors controlling expression for all but the LWS gene (Table 3).

3.5 Genetic mapping of regulatory loci

Single marker association tests for the opsin markers in the DC cross (Table 4A) found no association between opsin markers and opsin gene expression. This suggests that cis regulation is not important for this cross. However, this result could be limited by our only having 33 individuals available for this cross. Single marker tests with ten random markers did identify two trans-acting genomic regions associated with gene expression. The first region was identified with marker UNH871 on LG 13 and was associated with SWS1, SWS2B, and SWS2A expression. This marker was also linked to LWS expression and was nearly significant for RH2B expression as well. We selected ten additional markers on LG13 to test for association, though only four of these amplified and showed variation in this cross. Two of these markers were also associated with SWS expression, and all three had p values <0.05. We consider these three markers to be linked (see below) and so treat them as a single locus with regards Bonferroni correction for significance. A second region was identified with marker UNH2156 on LG4 and was associated with RH2A / LWS expression. This association was highly significant ($F_{2,28}=7.15$ p=0.003). We tested three additional markers on LG4, one of which (UNH2126) also showed significance. After sequential Bonferroni correction, only the region on LG4 is significantly associated with opsin expression.

Single marker associations were also tested in the TA cross for regions containing the opsin genes as well as the regions positively associated in the DC cross (Table 4B). We found positive association for a marker near red opsin with expression of SWS2B ($F_{2,44}=4.63$, p=0.015), SWS2A ($F_{2,44}=4.75$, p=0.013) and RH2B ($F_{2,44}=5.50$, p=0.007). The linkages with SWS2B and SWS2A were just beyond significance after sequential Bonferonni correction. We also found a marker near green opsin with positive association (RH2B $F_{2,44}=3.90$ p<0.03), though this was not significant after sequential Bonferroni correction. However, these loci suggest that cis regulation may be important for this cross. We also tested the two strongest trans acting regions identified in the DC cross and found strong association for markers on LG13 (p<0.004). However, there was no association with the marker on LG4. Therefore, both crosses showed strong association to a non-opsin region, LG4 for the DC cross and LG13 for the TA cross.

Figure 3 shows a summary of the results for the LG13 genomic region with several different maps. In Figure 3A, we show markers on LG 13 ordered by the tilapia reference map, noting the p values for association with SWS1 (DC cross) or SWS2b (TA cross) opsin expression. This figure also shows markers that were tried, but either did not amplify (NA), were not variable (NV) or were not scoreable (NS). In Figure 3B, we show the maps determined from recombination fraction between the three LG13 markers for each cross. There is strong evidence for an inversion in this region with a change in the Malawi species marker order, relative to the tilapia reference map. The Malawi marker order is 9700 times more likely than the tilapia marker order for the DC cross ($LOD_{Malawi} = 12.7$; $LOD_{tilapia}= 3.6$) and 17,700 times more likely for the TA cross ($LOD_{Malawi} = 24.4$; $LOD_{tilapia}= 14.6$). For both crosses, GM641 is closest to UNH871 and more distant from GM507. In addition, GM507 is even more distant from UNH871 in the DC cross relative to the TA cross, which perhaps explains why GM507 did not show association with opsin expression in this cross.

In Table 5, we quantify the fraction of variation in opsin expression which is explained by markers showing an association with opsin gene expression. This is determined from the

regression coefficient of phenotype (gene expression) on genotype. In the TA cross, the tightly linked markers significantly explain up to 25% of the variation. The fraction of variation explained in the DC cross is also significant with contributions up to 33% (LG4). However, we must note that it is possible that these contributions are overestimates as a result of the Beavis effect whereby QTL studies in small crosses yield larger relative contributions (Beavis, 1998).

Discussion

Opsin expression is labile

Our previous work suggests that opsin expression is highly variable. It can vary between adults of different species, as observed in cichlid species from Lake Malawi, as well as developmentally, as seen in the riverine tilapia, *O. niloticus* (Carleton et al., 2008). Opsin gene expression can even differ amongst closely related species within the same genus (Hofmann et al., 2009).

In this work, we examine the genetic architecture underlying opsin gene expression and find that expression is controlled by relatively few genes. Opsin expression in the F_2 for both crosses shows large variation, essentially covering the full range defined by the parentals. Castle Wright estimates suggest that the variance in gene expression is controlled by a small number of genetic factors, typically just 1-2 genes. One possible concern in making these estimates is the small number of parentals that we have for the DC cross and for *T. intermedius* in the TA cross. However, adding additional parental individuals (and increasing parental variance) would only decrease the estimated number of genes since the parental variance is subtracted in the numerator of the Castle-Wright estimator. Therefore, the estimate of a small number of genes is robust to our experimental sampling biases. This simple genetic architecture could facilitate evolutionary change in opsin expression between species as well as contribute to the evolutionary lability of opsin expression observed within the Malawi cichlid flock.

In addition, we have found that expression of certain opsin genes are highly correlated. We observed positive correlations between SWS1 and RH2B in the DC cross, and between SWS2B and RH2B in the TA cross. We also observed correlations between single and double cone wavelength, particularly in the DC cross. These correlations might explain why a few combinations of opsins are prevalent among cichlid species. Three opsin gene combinations are widespread in the three major cichlid radiations of Lakes Malawi, Victoria, and Tanganyika: UV (SWS1, RH2B, RH2A), violet (SWS2B, RH2B, RH2A) and blue (SWS2A, RH2A, LWS). These combinations also occur at different developmental stages, most notably in *O. niloticus* (Carleton et al., 2008). If certain genes are temporally correlated through development and correlated in different species, it seems possible that expression of the correlated opsin genes are controlled by the same transcription factor binding to a common promoter element for these genes. We are currently searching for commonalities in the promoter regions of correlated genes.

We estimate that there are 1-2 factors controlling expression of most of the opsin genes. Since there are 6 gene classes, this suggests 6-12 factors could be important. However, if transcription factors are indeed shared between genes, the number of genetic factors controlling opsin expression might be smaller. This commonality in transcription would further contribute to rapid changes in gene expression over evolutionary times, as well as evolution of a few key visual gene palettes.

Genetic basis of opsin regulation

We found evidence for both cis- and trans-acting factors regulating opsin expression. In the TA cross, there was one marker close to the SWS2A – SWS2B - LWS tandem array on LG5 that was associated with expression of RH2B (and nearly statistical for SWS2A and SWS2B expression). This marker explains 13-16% of the opsin variance. We did not find any cis-acting regions in the DC cross, though we had fewer individuals to test for association. There could be small cis regulatory effects in this cross that we did not have the power to detect.

We found strong evidence for trans-acting regulators of opsin gene expression. The TA cross showed associations with a region on LG13 which explained at least 20% of the variance. In addition, the DC cross showed an association with a region on LG4 which explained at least 30% of the variance. We have not yet fine mapped these regions. Once we have markers at the causative loci, they may explain even more of the variance.

The loci on LG4 and LG13 are trans to the opsin genes, which are located on LG17 (SWS1) and LG5 (SWS2-LWS and RH2 arrays). Therefore, they likely contain transcription factors or other regulatory molecules that modulate opsin expression. These trans acting factors may interact with multiple opsin promoters and could facilitate correlated change in several opsins.

Work in other species suggests a few trans-acting candidate genes which could be involved. Thyroid hormone is known to affect mouse cone opsin expression (Applebury et al., 2007, Ng et al., 2001) and both thyroid hormone and retinoic acid have been shown to alter opsin expression in salmon (Browman & Hawryshyn, 1994a, Browman & Hawryshyn, 1994b). The receptors of both of these molecules act as transcription factors through characteristic binding sites in gene promoters and often do so by forming heterodimers with each other (Glass, 1994). We hope to locate these receptors on the cichlid genetic map, using the forthcoming cichlid genome sequence and test for association with these important hormone receptors.

In the DC cross, we observed transgressive expression of the SWS2B and SWS2A genes, which either exceeded (SWS2B) or did not reach (SWS2A) the values of the two parental species. The SWS1 gene is on LG17 and SWS2a is on LG5. There may be some epistatic interactions between controlling loci which act to determine the balance between these genes. We did not see transgressive expression in the TA cross which involved SWS2B and SWS2A expression. These two genes are within 5 kb of each other on LG5 and so might be directly regulated by common factors.

Genetic architecture and comparison to other traits

The genetic architecture of several other cichlids traits have been determined for Lake Malawi rock dwellers. Several traits are Mendelian, based on a single gene. This includes tooth shape (Streelman et al., 2003b) and female orange-blotch color patterns (Roberts et al., 2009, Streelman et al., 2003a). Other traits are more genetically complex. Cichlid head and jaw morphology traits are controlled by between 5 and 10 genes (Albertson & Kocher, 2005, Albertson et al., 2003a, Albertson et al., 2003b). Blue-yellow coloration of different male color pattern elements were controlled by 4 to 7 loci (Barson et al., 2007). These traits were also influenced by dominance and epistasis. Our results for visual sensitivity fall in between the simple Mendelian and the complex traits, as expression of most of the opsins is controlled by a few factors. In contrast to jaw morphology or male color pattern, facile changes in opsin gene expression may enable visual sensitivity to shift, enabling rapid response to changes in the photic environment or the requirements of different foraging modes.

Models suggest that mating traits might be correlated with sensory traits (Endler, 1992). A mating trait is thereby selected to optimally stimulate the sensory pathway used to detect it. There is a close correlation between visual sensitivities and male mating coloration in Lake Victoria cichlids which is a result of changes in opsin sequence (Carleton et al., 2005, Seehausen et al., 2008). Species with longer wavelength visual sensitivity use longer wavelength mating colors. However, it is unclear whether vision drives color, color drives vision, or both are driven by the photic environment. Our previous studies suggest that evolution of visual sensitivity of Lake Malawi cichlids is controlled by differential gene expression. This causes large spectral shifts with finer scale shifts as a result of opsin sequence tuning (Hofmann et al., 2009). Gene expression is driven by fewer loci than male color, and is therefore potentially more labile. As a result, vision could respond more rapidly to selection from the environment, and therefore may drive the evolution of male color. Testing this hypothesis will require more explicit knowledge of the genetic basis of cichlid vision and color as well as knowledge of the evolutionary forces driving these traits.

Conclusions

Sensory evolution in cichlid fishes involves a number of different mechanisms. In previous work, we have demonstrated the importance of differential gene expression for producing large shifts in spectral sensitivity. In this work, we demonstrate that gene expression changes are likely controlled by just a few genetic factors. Since gene expression of certain genes is correlated, these controlling factors could drive several of the opsins. This could explain why there are just three common palette of genes expressed in wild cichlids. Here, we find strong evidence for two trans regulatory factors and a potential cis regulatory factor as well underlying opsin gene expression changes. Our previous work has shown that gene regulation, as well as coding sequence, play a role in cichlid visual sensitivity (Hofmann et al., 2009). With our finding that both cis and trans gene regulation are also important, it would seem that all possible mechanisms contribute to the spectral sensitivity of this key sensory system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Literature cited

Albertson RC, Kocher TD. Genetic architecture sets limits on transgressive segregation in hybrid cichlid fishes. Evolution Int J Org Evolution. 2005; 59:686–90.

Albertson RC, Streelman JT, Kocher TD. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. Proc Natl Acad Sci U S A. 2003a; 100:5252–7. [PubMed: 12704237]

Albertson RC, Streelman JT, Kocher TD. Genetic basis of adaptive shape differences in the cichlid head. J Hered. 2003b; 94:291–301. [PubMed: 12920100]

Applebury ML, Farhangfar F, Glosmann M, Hashimoto K, Kage K, Robbins JT, Shibusawa N, Wondisford FE, Zhang H. Transient expression of thyroid hormone nuclear receptor TRbeta2 sets S opsin patterning during cone photoreceptor genesis. Dev Dyn. 2007; 236:1203–12. [PubMed: 17436273]

Barson NJ, Knight ME, Turner GF. The genetic architecture of male colour differences between a sympatric Lake Malawi cichlid species pair. J Evol Biol. 2007; 20:45–53. [PubMed: 17209998]

Beavis, WD. QTL Analyses: Power, Precision, and Accuracy. In: Paterson, AH., editor. Molecular Dissection of Complex Traits. CRC Press; Boca Raton, FL: 1998. p. 145-162.

Boughman JW. How sensory drive can promote speciation. Trends in Ecology and Evolution. 2002; 17:571–7.

Browman H, Hawryshyn C. Retinoic Acid Modulates Retinal Development in the Juveniles of a Teleost Fish. J Exp Biol. 1994a; 193:191–207. [PubMed: 9317609]

Browman HI, Hawryshyn CW. Thyroxine induces a precocial loss of ultraviolet photosensitivity in rainbow trout (Oncorhynchus mykiss, Teleostei). Vision Res. 1992; 32:2303–12. [PubMed: 1288007]

Browman HI, Hawryshyn CW. The developmental trajectory of ultraviolet photosensitivity in rainbow trout is altered by thyroxine. Vision Res. 1994b; 34:1397–406. [PubMed: 8023449]

Carleton KL. Cichlid fish visual systems: mechanisms of spectral tuning. Integrative Zoology. 2009; 4:75–86. [PubMed: 21392278]

Carleton KL, Kocher TD. Cone opsin genes of african cichlid fishes: tuning spectral sensitivity by differential gene expression. Mol Biol Evol. 2001; 18:1540–50. [PubMed: 11470845]

Carleton KL, Parry JW, Bowmaker JK, Hunt DM, Seehausen O. Colour vision and speciation in Lake Victoria cichlids of the genus Pundamilia. Mol Ecol. 2005; 14:4341–53. [PubMed: 16313597]

Carleton KL, Spady TC, Streelman JT, Kidd MR, McFarland WN, Loew ER. Visual sensitivities tuned by heterochronic shifts in opsin gene expression. BMC Biol. 2008; 6:22. [PubMed: 18500997]

Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008; 134:25–36. [PubMed: 18614008]

Dann SG, Ted Allison W, Veldhoen K, Johnson T, Hawryshyn CW. Chromatin immunoprecipitation assay on the rainbow trout opsin proximal promoters illustrates binding of NF-kappaB and c-jun to the SWS1 promoter in the retina. Exp Eye Res. 2004; 78:1015–24. [PubMed: 15051482]

Endler JA. Signals, signal conditions and the direction of evolution. Am Nat. 1992; 139:S125–S153.

Fernald RD. Chromatic organization of a cichlid fish retina. Vision Res. 1981; 21:1749–1753. [PubMed: 7336611]

Fraser GJ, Bloomquist RF, Streelman JT. A periodic pattern generator for dental diversity. BMC Biol. 2008; 6:32. [PubMed: 18625062]

Fraser GJ, Hulsey CD, Bloomquist RF, Uyesugi K, Manley NR, Streelman JT. An ancient gene network is co-opted for teeth on old and new jaws. PLoS Biol. 2009; 7:e31. [PubMed: 19215146]

Fryer, G.; Iles, TD. The cichlid fishes of the great lakes of Africa: Their biology and evolution. TFH Publications; Neptune City, NJ: 1972.

Glass CK. Differential Recognition of Target Genes by Nuclear Receptor Monomers, Dimers, and Heterodimers. Endocrine Reviews. 1994; 15:391–407. [PubMed: 8076589]

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature. 2005; 433:481–7. [PubMed: 15690032]

Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika. 1988; 75:800–802.

Hoekstra HE, Coyne JA. The locus of evolution: evo devo and the genetics of adaptation. Evolution Int J Org Evolution. 2007; 61:995–1016.

Hofmann CM, O'Quin KE, Marshall NJ, Cronin TC, Seehausen O, Carleton KL. The eyes have it: Regulatory and structural changes both underlie cichlid visual pigment diversity. PLoS Biol. 2009; 7:e1000266. [PubMed: 20027211]

Jones I, Srinivas M, Ng L, Forrest D. The thyroid hormone receptor beta gene: structure and functions in the brain and sensory systems. Thyroid. 2003; 13:1057–68. [PubMed: 14651789]

Jordan R, Kellogg K, Howe D, Juanes F, Stauffer JR, Loew ER. Photopigment spectral absorbance of Lake Malawi cichlids. J Fish Biology. 2006; 68:1291–9.

Kelley MW, Turner JK, Reh TA. Ligands of steroid/thyroid receptors induce cone photoreceptors in vertebrate retina. Development. 1995; 121:3777–85. [PubMed: 8582287]

King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. Science. 1975; 188:107–16. [PubMed: 1090005]

Kocher TD. Adaptive evolution and explosive speciation: the cichlid fish model. Nat Rev Genet. 2004; 5:288–98. [PubMed: 15131652]

Konings, A. Malawi cichlids in their natural habitat, 4th ed. 2nd. Cichlid Press; El Paso, TX: 2007.

Lande R. The minimum number of genes contributing to quantitative variation between and within populations. Genetics. 1981; 99:541–53. [PubMed: 7343418]

Lee BY, Lee WJ, Streelman JT, Carleton KL, Howe AE, Hulata G, Slettan A, Stern JE, Terai Y, Kocher T. A second generation genetic linkage map of tilapia (Oreochromis spp.). Genetics. 2005; 170:237–244. [PubMed: 15716505]

Liem, KF. Functional morphology. In: Keenleyside, MHA., editor. Cichlid fishes: behaviour, ecology and evolution. Chapman and Hall; New York: 1991. p. 129-150.

Liu, BH. Statistical Genomics: Linkage, Mapping, and QTL Analysis. CRC Press; New York: 1998.

Luo W, Williams J, Smallwood PM, Touchman JW, Roman LM, Nathans J. Proximal and distal sequences control UV cone pigment gene expression in transgenic zebrafish. J Biol Chem. 2004; 279:19286–93. [PubMed: 14966125]

Lynch, M.; Walsh, B. Genetics and analysis of quantitative traits. Sinauer; Sunderland, MA: 1998.

Marchand O, Duffraisse M, Triqueneaux G, Safi R, Laudet V. Molecular cloning and developmental expression patterns of thyroid hormone receptors and T3 target genes in the turbot (Scophtalmus maximus) during post-embryonic development. Gen Comp Endocrinol. 2004; 135:345–57. [PubMed: 14723886]

Neitz M, Neitz J. The uncommon retina of the common house mouse. Trends Neurosci. 2001; 24:248–50. [PubMed: 11311361]

Ng L, Hurley JB, Dierks B, Srinivas M, Salto C, Vennstrom B, Reh TA, Forrest D. A thyroid hormone receptor that is required for the development of green cone photoreceptors. Nat Genet. 2001; 27:94–8. [PubMed: 11138006]

Ott, J. Analysis of Human Genetic LInkage. The Johns Hopkins University Press; Baltimore: 1991.

Paris M, Laudet V. The history of a developmental stage: metamorphosis in chordates. Genesis. 2008; 46:657–72. [PubMed: 18932261]

Parry JW, Carleton KL, Spady T, Carboo A, Hunt DM, Bowmaker JK. Mix and match color vision: tuning spectral sensitivity by differential opsin gene expression in Lake Malawi cichlids. Curr Biol. 2005; 15:1734–9. [PubMed: 16213819]

Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh SD, True JR, Carroll SB. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature. 2006; 440:1050–3. [PubMed: 16625197]

Roberts R, Ser J, Kocher TD. Genetic basis of a sexual conflict in Lake Malawi cichlids. Science. 2009 e-pub.

Ryan MJ, Rand AS. Sexual selection and signal evolution: The ghost of biases past. Philos Trans R Soc Lond B Biol Sci. 1993; 340:187–195.

Salzburger W. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol Ecol. 2009; 18:169–85. [PubMed: 18992003]

Schmitt EA, Hyatt GA, Dowling JE. Erratum: Temporal and spatial patterns of opsin gene expression in the zebrafish (Danio rerio): corrections with additions. Vis Neurosci. 1999; 16:601–5. [PubMed: 10349978]

Seehausen, O. Lake Victoria rock cichlids. Verduijn Cichlids; Germany: 1996.

Seehausen O. African cichlid fish: a model system in adaptive radiation research. Proc Biol Sci. 2006; 273:1987–98. [PubMed: 16846905]

Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HD, Miyagi R, van der Sluijs I, Schneider MV, Maan ME, Tachida H, Imai H, Okada N. Speciation through sensory drive in cichlid fish. Nature. 2008; 455:620–6. [PubMed: 18833272]

Spady TC, Parry JW, Robinson PR, Hunt DM, Bowmaker JK, Carleton KL. Evolution of the cichlid visual palette through ontogenetic subfunctionalization of the opsin gene arrays. Mol Biol Evol. 2006; 23:1538–47. [PubMed: 16720697]

Streelman JT, Albertson RC, Kocher TD. Genome mapping of the orange blotch colour pattern in cichlid fishes. Mol Ecol. 2003a; 12:2465–71. [PubMed: 12919484]

Streelman JT, Webb JF, Albertson RC, Kocher TD. The cusp of evolution and development: a model of cichlid tooth shape diversity. Evol Dev. 2003b; 5:600–8. [PubMed: 14984042]

Terai Y, Seehausen O, Sasaki T, Takahashi K, Mizoiri S, Sugawara T, Sato T, Watanabe M, Konijnendijk N, Mrosso HD, Tachida H, Imai H, Shichida Y, Okada N. Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. PLoS Biol. 2006; 4:e433. [PubMed: 17147472]

Tsujimura T, Chinen A, Kawamura S. Identification of a locus control region for quadruplicated green-sensitive opsin genes in zebrafish. Proc Natl Acad Sci U S A. 2007; 104:12813–8. [PubMed: 17646658]

van der Meer HJ, Bowmaker JK. Interspecific variation of photoreceptors in four co-existing haplochromine cichlid fishes. Brain Behav Evol. 1995; 45:232–40. [PubMed: 7620872]

Wald G. Carotenoids and the Visual Cycle. J Gen Physiol. 1935; 19:351–371. [PubMed: 19872932]

Wray GA. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. 2007; 8:206–16. [PubMed: 17304246]

Wright, S. Evolution and the Genetics of Populations. University of Chicago Press; Chicago, IL: 1968.

Zeng ZB. Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improved method. Genetics. 1992; 131:987–1001. [PubMed: 1325390]

Zeng ZB, Houle D, Cockerham CC. How informative is Wright's estimator of the number of genes affecting a quantitative character? Genetics. 1990; 126:235–47. [PubMed: 2227383]
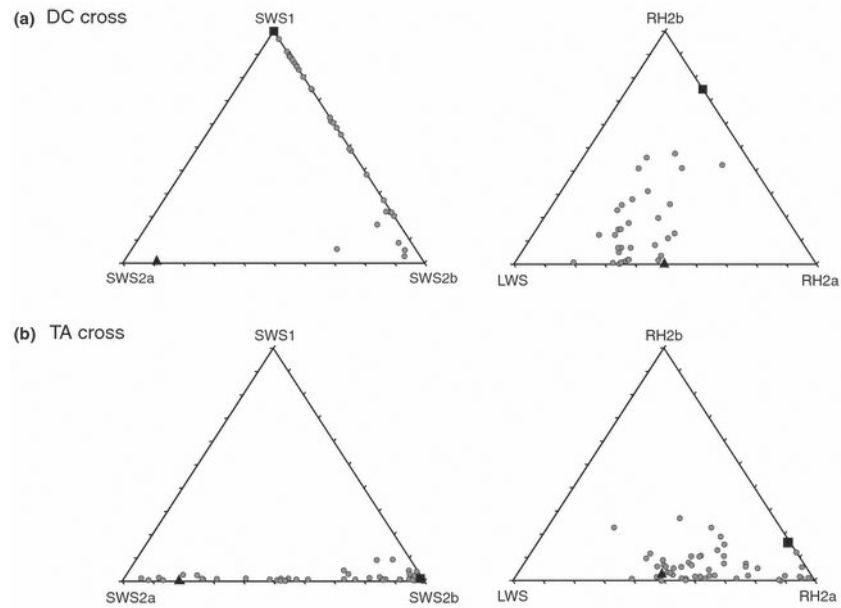
Figure 1.
Triangle plots of single (left) and double (right) cones for the DC and TA crosses showing parental and F2 (λ) gene expression. For the DC cross, σ = *Dimidiochromis* and ν =*Copadichromis*. For the TA cross σ = *Tramitichromis*, ν = *Aulonocara*.
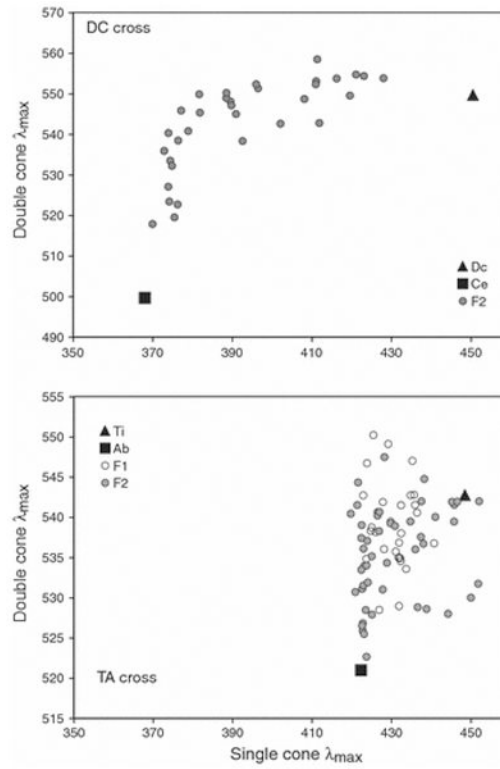
Figure 2.
A plot of the average single versus average double cone wavelengths for the DC and TA crosses including parentals and F_2. For the TA cross, F_1 are also shown. Average wavelengths are calculated based on individual gene expression using equations 3a and 3b.
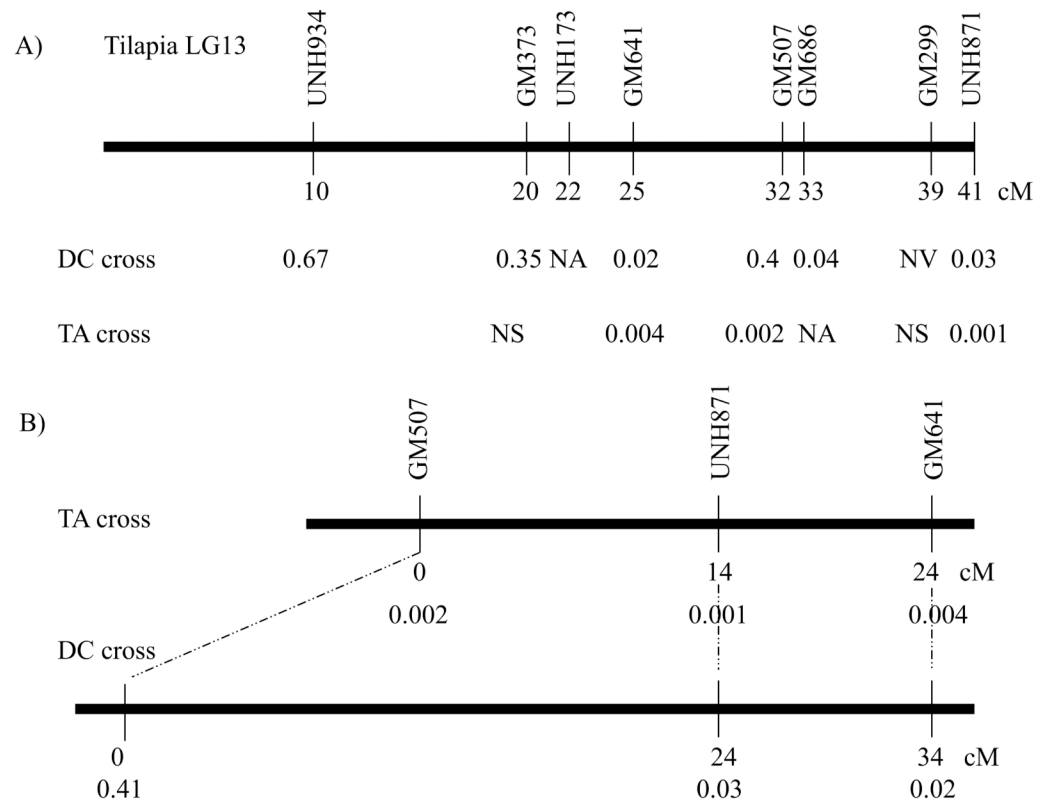
Figure 3.
Genomic region on LG13 showing association with opsin expression. In A) markers are given in the order of the tilapia LG13 map. Markers are labeled with p values for association with SWS1 (DC cross) or SWS2b (TA cross) expression. In B) maps are based on recombination events for the actual TA or DC cross individuals for scored markers.

Table 1

Genotyping primers for markers close to the opsin genes. Primer sequences and location relative to the opsins genes are listed.

Marker	Cross	Location and Primer sequences
UVusat (TCAA)$_4$	Both	Third intron of SWS1
		F: GCGCGGAATTCACATCCCTGAAAGTCTGGGC
		R: GCGCGCAAGCTTCTGTTTGTTCATGAAGACGTAGA
Green2 (GTT)$_{13}$	DC	4.9 kb from end of RH2B and 3.8 kb from begin of RH2Aα
		F: TGCCTTTGGTGAGTCTTTTT
		R: CCACATTGGAATAAAGTTCATGTT
Green1usat2 (CA)$_{43}$	TA	11.4 kb from end of RH2Aβ
		F: CCACAATTACACTGCGAAGG
		R: TGAATGGTGTGCCTGTTGTT
Redblue4 (CT)$_{15}$	DC	22 kb from end of LWS
		F: AGTCACAAGCCATAAATACTTGAACA
		R: AATATATTACATTTTCGGTCCGTCTC
Redblue3 (CA)$_{30}$	TA	7.6 kb from end of LWS
		F: TGTTCCATCATTCAACTTCATCA
		R: AAGAACACCATGGCAGGATT

Table 2

Gene correlations. The F statistic and p value for the regressions are listed for each opsin gene combination (*p<0.01, **p<10⁻⁴). If the p value is statistically significant, the regression coefficient is also listed.

A. DC cross.

	RH2A	RH2B	SWS2A	SWS2B	SWS1
LWS	$F_{1,31}=2.7$ p=0.11	$F_{1,31}=61.3$ p=7.8e-9** -0.79	$F_{1,31}=1.4$ p=0.24	$F_{1,31}=8.2$ p=0.007 0.70	$F_{1,31}=16.3$ p=3.3e-4** -1.087
RH2A		$F_{1,31}=1.3$ p=0.26	$F_{1,31}=0.05$ p=0.83	$F_{1,31}=0.01$ p=0.92	$F_{1,31}=1.3$ p=0.27
RH2B			$F_{1,31}=3.1$ p=0.09	$F_{1,31}=29.5$ p=6.2e-6** -1.10	$F_{1,31}=33.5$ p=2.3e-6** 1.38
SWS2A				$F_{1,31}=4.0$ p=0.053	$F_{1,31}=8.2$ p=0.007* -0.09
SWS2B					$F_{1,31}=60.2$ p=9.3e-9** -0.99

B. TA cross

	RH2A	RH2B	SWS2A	SWS2B	SWS1
LWS	$F_{1,48}=128.2$ p=3.7e-15*** -0.77	$F_{1,48}=0.01$ p=0.94	$F_{1,48}=0.003$ p=0.96	$F_{1,48}=0.4$ p=0.55	$F_{1,48}=3.6$ p=0.06
RH2A		$F_{1,48}=11.7$ p=0.001* -1.36	$F_{1,48}=4.8$ p=0.03 0.71	$F_{1,48}=6.9$ p=0.01 -0.61	$F_{1,48}=3.1$ p=0.08
RH2B			$F_{1,48}=30.0$ p=1.6e-6** -0.48	$F_{1,48}=39.4$ p=9.4e-8** 0.37	$F_{1,48}=0.05$ p=0.83
SWS2A				$F_{1,48}=161.3$ p=5.8e-17** -0.64	$F_{1,48}=2.1$ p=0.16
SWS2B					$F_{1,48}=0.8$ p=0.37

Table 3

Castle Wright estimates of number of factors (n_e) controlling expression of each of the opsins. For the TA cross, we also list the variance of n_e (Var n_e), dominance (d) and n_e estimated with the effecs of dominance ($n_{e, dom}$).

A. DC cross

	SWS1	SWS2B	SWS2A	RH2B	RH2A	LWS
Dc mean	1.25	10.67	88.09	0.45	38.16	38.51
Ce mean	99.98	0.01	0.01	46.73	15.48	0.00
Var Dc	0.00	0.14	16.30	0.00	14.72	0.29
Var Ce	3.61	0.00	0.00	1.69	0.36	0.00
Var F_2	958.03	835.65	22.43	141.35	44.68	132.68
Var F_1*	3.61	0.14	16.3	1.69	14.72	0.29
n_e	1.28	0.02	157.7	1.92	2.08	1.40

* estimated as the larger of Var Dc or Var Ce

B. TA cross

	SWS2B	SWS2A	RH2B	RH2A	LWS
Ti mean	18.47	80.85	3.19	47.78	49.03
Ab mean	98.06	0.77	16.37	82.92	0.71
Var Ti	5.15	7.14	0.14	0.65	0.19
Var Ab	4.41	0.28	3.03	4.65	1.38
Var F_2	885.28	899.54	40.53	312.23	263.07
Var F_1	184.79	176.27	15.01	162.89	186.38
n_e	1.13	1.11	0.83	1.03	3.80
Var(n_e)	0.09	0.08	0.13	0.28	12.3
F_1 mean	70.41	27.31	5.96	52.89	41.15
Dominance, d	0.15	0.17	-0.29	-0.35	-0.34
$n_{e, dom}$	1.19	1.14	0.99	1.4	4.54

Table 4

Test for single marker association between markers at the opsins or elsewhere in the genome and gene expression for each opsin. Linkage group (and position in cM) is taken from the tilapia map (Lee et al 2005). The relative marker positions on LG4 are taken from the mbuna map of Albertson et al 2003 relative to the known position of UNH911 on the tilapia map. For each marker, we list the number of individuals that were successfully genotyped, the degrees of freedom for the F statistic, and then for each gene we list the F statistic and p value. Markers which are statistically significant after sequential Bonferroni correction are shown in bold. (p values are given in Supplementary Table 1).

A. DC cross:

Marker	Linkage group (cM)	# ind genotyped	d.f.	F statistic ** p<0.01					
				fSWS1	fSWS2B	fSWS2A	fRH2B	fRH2A	fLWS
Green opsin	5 (18)	26	3, 22	0.699	0.658	0.789	0.750	1.392	0.084
Red-blue opsin	5 (48)	29	2, 26	0.066	0.070	1.160	0.628	0.524	0.502
UV opsin	17	28	2, 25	0.549	0.610	0.113	0.009	2.394	0.631
UNH2008	2	28	2, 25	1.769	1.627	2.935	1.099	1.986	0.403
UNH2016	4 (16)	31	2, 28	0.061	0.046	0.530	0.746	1.461	0.340
UNH2126	4 (19)	29	2, 26	0.481	1.250	3.004	0.627	5.988	3.092
UNH2156	4 (20)	31	2, 28	0.013	0.015	0.424	0.748	**7.152****	4.274
UNH911	4 (22)	NS							
UNH2003	6	29	2, 26	0.400	0.411	0.094	0.051	0.815	0.055
UNH2164	6	31	2, 28	0.015	0.090	1.237	0.442	0.943	1.332
UNH2153	7	32	1, 30	1.267	1.097	0.867	0.088	0.005	0.060
UNH129	8	32	3, 28	2.073	2.020	2.396	2.143	0.849	1.541
UNH875	9	29	2, 26	0.001	0.008	0.683	0.080	0.087	0.010
UNH2084	12	32	2, 29	1.715	1.579	2.728	1.878	0.204	2.572
UNH934	13 (10)	31	3, 27	0.525	0.393	1.046	0.610	0.118	0.381
GM373	13 (20)	31	1, 29	0.909	0.699	1.209	0.724	0.004	0.722
UNH173	13 (22)	NA		NV					
GM641	13 (25)	32	2, 29	4.342	3.800	3.460	2.340	0.436	1.684
UNH997	13 (31)	NV							
GM262	13 (32)	32	1, 30	1.033	0.724	2.033	2.307	0.920	0.800
GM507	13 (32)	30	2, 27	0.916	0.883	0.823	0.916	1.969	2.056
UNH954	13 (32)			NV					
GM686	13 (33)	28	1, 26	4.779	5.527	0.047	2.574	0.320	1.405

A. DC cross:

Marker	Linkage group (cM)	# ind genotyped	d.f.	F statistic ** p<0.01					
				fSWS1	fSWS2B	fSWS2A	fRH2B	fRH2A	fLWS
UNH1008	13 (34)			NA					
GM299	13 (39)			NV					
UNH871	13 (41)	29	2, 26	3.883	3.041	3.703	3.293	0.288	3.587
UNH2166	15	30	3, 27	0.274	0.316	0.638	1.468	0.955	3.072

NA = no amplification

NS = not scorable

NV = not variable in cross

B. TA cross

Marker	Linkage group (cM)	# ind genotyped	d.f.	F statistic *p<0.05; ** p<0.01				
				SWS2B	SWS2A	RH2B	RH2A	LWS
Green opsin	5	47	2, 44	0.500	0.439	3.901	0.633	0.614
Red-blue opsin	5	47	2, 44	4.637	4.756	5.503*	2.033	0.529
UV opsin	17	44	2, 41	1.644	1.729	1.182	0.021	0.063
UNH2156	4	45	2, 42	0.244	0.245	0.504	0.838	1.099
GM373	13 (2)	NS	2, 44					
GM641	13 (25)	47	2, 44	6.415**	6.338**	0.750	0.415	0.766
GM262	13 (32)	NV						
GM507	13 (32)	47	2, 44	6.998**	7.520**	0.986	1.299	0.765
GM686	13 (33)	NA						
GM299	13 (39)	NS						
UNH871	13 (41)	47	2, 44	7.638**	7.666**	0.354	0.846	0.999

NA = no amplification

NV = no variation

NS = not scorable

Table 5

Regression statistics for markers with positive association, including regression coefficient (R^2), F statistic and p value.

A. DC cross

Marker	Linkage Grp	fSWS1	fSWS2B	fSWS2A	fRH2B	fRH2A	fLWS
	4					0.25	0.19
						$F_{1,27}=8.97$	$F_{1,27}=6.28$
UNH2126						0.006**	0.018*
	4					0.33	0.18
						$F_{1,29}=14.51$	$F_{1,29}=6.56$
UNH2156						0.0007**	0.018*
	13	0.001	0.0002	0.07			
		$F_{1,30}=0.027$	$F_{1,30}=0.005$	$F_{1,30}=2.32$			
GM641		0.87	0.95	0.14			
	13	0.16	0.18				
		$F_{1,26}=4.78$	$F_{1,26}=5.53$				
GM686		0.038*	0.0268				
	13	0.03		0.07			0.002
		$F_{1,27}=0.83$		$F_{1,27}=2.2$			$F_{1,27}=0.05$
UNH871		0.37		0.15			0.82

B. TA cross

Marker	Linkage group	SWS2B	SWS2A	RH2B	RH2A	LWS
		0.16	0.16	0.13		
		$F_{45,1}=8.34$	$F_{45,1}=8.70$	$F_{45,1}=7.04$		
Red opsin	5	0.006**	0.005**	0.011*		
		0.22	0.22			
		$F_{45,1}=12.99$	$F_{45,1}=12.66$			
GM641	13	0.0008**	0.0009**			
		0.24	0.25			
		$F_{45,1}=14.2$	$F_{45,1}=15.2$			
GM507	13	0.0005**	0.0003**			
		0.25	0.25			

B. TA cross

Marker	Linkage group	SWS2B	SWS2A	RH2B	RH2A	LWS
		$F_{45,1}=14.8$	$F_{45,1}=14.6$			
UNH871	13	0.0004	0.0004			