ABSTRACT

Dissertation Title: MECHANISMS CONTRIBUTING TO OPSIN EXPRESSION DIVERGENCE IN THE VISUAL SYSTEM OF AFRICAN CICHLIDS

Sri Pratima Nandamuri, Doctor of Philosophy
2018

Dissertation directed by: Professor Karen L Carleton,
Department of Biology

Vision is an important sensory modality, guiding fundamental tasks such as foraging, escaping from predators, identification of conspecifics and selection of mates. As such, animals exhibit a wide variety of adaptations to spectrally tune their visual systems to closely match the environment. This extensive variation in visual system tuning is achieved both via genetic changes as well as environmental induced plasticity.

Lacustrine cichlids of East Africa are famous for their expedited adaptive radiations. Cichlids in these rift lakes inhabit a diverse range of light environments, from the murky red-shifted waters of Lake Victoria to clear Lake Malawi. Consequently, African cichlids have some of the most diverse visual systems among vertebrates on the planet, with species

expressing different combinations of seven cone opsin genes. This differential expression is under genetic control and leads to drastic differences in the visual sensitivities between closely related species. Moreover, cichlid species often exhibit plastic changes in opsin expression due to alterations in environmental light conditions. The diversity of genetically determined visual palettes and the variation in expression due to plastic changes offers an excellent opportunity to study the proximate mechanisms governing opsin expression divergence in this group.

Utilizing a hybrid cross between two species varying in opsin expression, we show that divergent expression of cone opsins is regulated by multiple quantitative trait loci (QTLs). Most of these QTLs are located in *trans* to the opsins, with the exception of one QTL in *cis* to the *SWS1* opsin. Further fine-mapping of the *cis*-QTL revealed a deletion in the promoter of the *SWS1* gene that is associated with a decrease in its expression. Additionally, performing two reciprocal experiments, we show that adult cichlids from Lake Malawi show rapid and reversible plastic changes in opsin expression due to differences in lighting conditions. These studies show that both predetermined genetic factors and environmental inputs contribute to opsin expression divergence in cichlids. These factors enable short term and ultimately long-term adaptation to changing habitats, facilitating the survival and perhaps speciation of these fantastic fishes.

MECHANISMS CONTRIBUTING TO OPSIN EXPRESSION DIVERGENCE IN THE VISUAL SYSTEM OF AFRICAN CICHLIDS

by

Sri Pratima Nandamuri

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Karen Carleton, Chair
Associate Professor Alexa Bely
Professor Eric Haag
Professor Thomas Kocher
Professor Leslie Pick

Foreword

Chapter 1: Figure 1.5 (top panel) is used with permission from Kelly O' Quin.

Chapter 2: Previously published under: Nandamuri SP, Dalton B, Carleton K: Determination of the genetic architecture underlying short wavelength sensitivity in Lake Malawi cichlids. *Journal of Heredity* 2017, 108(4):379-390.

I made the experimental cross, designed the study, performed the quantitative analysis, microsatellite mapping and plasticity experiments, and drafted the manuscript. Brian Dalton performed the in-situ experiments. Karen Carleton conceptualized the study and critically reviewed the article.

Chapter 3: In Review under: Nandamuri SP, Conte M, Carleton K: Multiple *trans* factors and one *cis*-regulatory deletion contribute to differential expression of opsins in African cichlids. *BMC Genomics*: 2018.

I generated the hybrid cross, performed the QTL experiment and analyses, and drafted the manuscript. Matthew Conte performed the bioinformatic analysis pertaining to the study and reviewed the manuscript. Karen Carleton conceptualized the study and critically reviewed the article.

Chapter 4: Previously published under: Nandamuri SP, Yourick MR, Carleton KL: Adult plasticity in African cichlids: Rapid changes in opsin expression in response to environmental light differences. *Mol Ecol* 2017, 26(21):6036-6052.

I designed the Natural to Narrow experiment, and Miranda Yourick designed the Narrow to Broad experiment. I and Miranda Yourick performed analyses and drafted the manuscript. Karen Carleton conceptualized the study and critically reviewed the article.

Chapter 5: Jane Schulte fine-mapped the QTL and identified candidate genes. Benjamin Sandkam was responsible for grandparent genome assembly and Matt Conte evaluated the repeat sequences bioinformatically. Karen Carleton conceptualized the study and critically reviewed the article. I performed all experiments pertaining to candidate gene evaluation and drafted the chapter.

Dedication

To my grandmother Lalitha Pasupaleti, I owe it all to you!

Thank you.

Acknowledgements

First and foremost, I want to thank my advisor Karen Carleton. Karen provided me with every bit of guidance and expertise to make my PhD experience productive and stimulating. She gave me the freedom to choose what projects I wanted to work on, while continuing to contribute valuable feedback and encouragement. Outside the lab, I am especially grateful for the personal rapport we have forged over the years. Always patient and ready with a listening ear, she is my go to person for advice on matters related to my professional career and personal life. In Karen, I found a mentor for life.

I wish to thank my committee members Eric Haag, Leslie Pick and Alexa Bely for their time and constructive feedback throughout this entire PhD process. I want to particularly acknowledge Tom Kocher for all the thoughtful insights into my work.

My deepest appreciation goes to the cichlid crew who stuck it out in grad school with me: Daniel Escobar-Camacho, Miranda Yourick, Matt Conte, Will Gammerdinger and Betsy Clark. They made grad school fun and at times bearable. I like to thank Marissa Schoepp for logging long hours in the fishroom, helping me collect the broods necessary for my experiments, and to Ben Sandkam for all the valuable suggestions with experiments and writing. I would also like to thank past lab members Jane Schulte, Brian Dalton, Conor O' Brien, Kelly O' Quin and Claire O' Quin for intellectual support.

My gratitude extends to my family for their love and unconditional support. I want to thank my mom Indira Rani for leading by example. It couldn't have been easy doing a PhD while balancing the rigors of a full-time job and raising two kids, but my mom showed me that if she could do it, I might as well try. My dad Venkateswara Rao for championing my cause since day one, no questions asked. My sister Pranathi for being my cheerleader. And most of all for my loving, patient and encouraging husband Prem, without whom this work would never have come to fruition.
Thank you.

Sri Pratima Nandamuri
University of Maryland
2018.

Table of Contents

Foreword ii

Dedication iii

Acknowledgements iv

Table of Contents v

List of Tables vii

List of Figures viii

Chapter1: Introduction 1

1.1 INTRODUCTION 2
1.2 VISUAL SYSTEMS 4
1.3 VISUAL SYSTEMS IN AFRICAN CICHLIDS 5
1.4 IMPORTANCE OF VISION TO CICHLIDS 8
1.5 GOALS OF RESEARCH 10

Chapter 2: Determination of the genetic architecture underlying short wavelength sensitivity in Lake Malawi cichlids. 14
2.1 ABSTRACT 15
2.2 BACKGROUND 16
2.3 MATERIALS AND METHODS 21
2.4 RESULTS 27
2.5 DISCUSSION 33
2.6 CONCLUSION 44

Chapter 3: Multiple *trans* factors and one *cis*-regulatory deletion contribute to differential expression of opsins in African cichlids. 46
3.1 ABSTRACT 47
3.2 BACKGROUND 48
3.3 METHODS 53
3.4 RESULTS 60
3.5 DISCUSSION 75
3.6 CONCLUSION 84

Chapter 4: Adult plasticity in African Cichlids: Rapid changes in opsin expression in response to environmental light differences. 86
4.1 ABSTRACT 87
4.2 BACKGROUND 88
4.3 MATERIALS AND METHODS 93
4.4 RESULTS 100
4.5 DISCUSSION 106
4.6 CONCLUSION 116

Chapter 5: A second gene contributing to the differential expression of *SWS2A* expression in Lake Malawi cichlids. 124
5.1 ABSTRACT 125

5.2 BACKGROUND 126
5.3 MATERIALS AND METHODS 129
5.4 RESULTS AND DISCUSSION 135
5.5 DISCUSSION 147
5.6 CONCLUSION 150
Chapter 6: Synthesis 153
6.1 INTRODUCTION 154
6.2 MAIN INFERENCES 154
6.3 FUTURE DIRECTIONS 157
REFERENCES 160

List of Tables

Table 2.1: Opsin-specific Taqman primer and probes used for RT-PCR……...28
Table 2.2: Genotyping primers for markers located in close proximity to the opsins……..29
Table 2.3: Means and variances (in parenthesis) of relative opsin expression in the F0, F1 and F2 generations of the hybrid cross……30
Table 2.4: Opsin gene correlations and p-values (in parenthesis) in F2 hybrids……....30
Table 2.5: Castle-Wright estimates of the number of factors (N_e) controlling the expression of each opsin……..32
Table 3.1: Sequencing primers for the promoter of the *SWS1* opsin gene……...59
Table 3.2: Summary of QTL for single cone opsin expression in the MA cross……65
Table 3.3: Summary of QTL for double cone opsin expression in the MA cross……65
Table 3.4: Putative Candidate genes for the differential expression of opsins in all QTLs except LG17……...72
Table 3.5: Potential targets of miR-729 and transcription factors (TFs) with putative binding sites in the CNE……74
Table 4.1: Post-hoc Tukey's comparisons for all the opsins for Natural to Narrow experiment……...100
Table 4.2: Post-hoc Tukey's comparisons for all the opsins for Narrow to Broad experiment……105
Table 4.3: Quantum catches of the opsins for different lighting environments……107
Table 5.1: Taqman primer and probe sequences for the candidates on LG5 and housekeeping gene *GNAT2*……..…132
Table 5.2: Sequencing primers for *MITFA* gene and promoter.....137
Table 5.3: List of genes in the 95% CI on LG5 that were differentially expressed between the F0 of the cross and correlated with *SWS2A* expression……...138
Table 5.4: Two-way genotype table with average expression of *SWS2A* for each genotype at both *MITFA* and *RX1* loci in the F2 hybrids……..144
Table 5.5: Two-way genotype table with average expression of *SWS2A* for each genotype at both *MITFA* and *RX1* loci in the Lake Malawi species panel……145

List of Figures

Fig. 1.1: Genetic and environmental inputs shape phenotypes……2
Fig. 1.2: Cichlid cone opsin genes……6
Fig. 1.3: Visual pigment templates in three Lake Malawi species……7
Fig. 1.4: Developmental plasticity in *SWS1* opsin expression……..9
Fig. 1.5: QTL study for cone opsin expression in the *Tramitochromis X Aulonocara* cross……...12
Fig. 2.1: Variation in opsin expression between adult wild-caught and lab-reared adult cichlid species……....37
Fig. 2.2: Boxplots of opsin expression from the parental, F1 and F2 generations in the MA cross……..…….39
Fig. 2.3: *In-situ* hybridization showing coexpression of *SWS1* and *SWS2B* opsins in the retinas of F2 hybrids……..……40
Fig. 2.4: *Cis*-variation for *SWS1* expression……...42
Fig. 3.1: Linkage map constructed from 157 F2 hybrids of an intergeneric cross between two species from Lake Malawi……...……66
Fig. 3.2: Quantitative trait loci (QTL) mapping in the MA cross……67
Fig. 3.3: QTL mapping for *SWS1* and *SWS2B* opsins……68
Fig 3.4: Mean opsin expression among genotypic classes at each single cone QTL…...…69
Fig. 3.5: QTL mapping for *RH2B*, *RH2A* and LWS……....70
Fig. 3.6: Mean double opsin expression among genotypic classes at each QTL……....71
Fig. 3.7: Comparison of coding sequences between *A. baenschi* and *M. mbenji*…….....75
Fig. 3.8: *Cis*-regulatory mutation in the *SWS1* opsin regulates its expression……76
Fig. 3.9: Phylogenetic footprinting of the *SWS1* promoter in teleosts……79
Fig. 4.1: Scheme for the experimental light conditions and sampling time points……..98
Fig. 4.2: Natural to Narrow opsin expression for *SWS1* and *SWS2B* opsins……..106
Fig. 4.3: Natural to Narrow opsin expression for *SWS2A, RH2B, RH2A& LWS* opsins…117
Fig. 4.4: Narrow to Broad opsin expression for *SWS1*, *SWS2B* and *LWS* opsins……...118
Fig. 4.5: Narrow to Broad opsin expression for *SWS2A*, *RH2B* and *RH2A* opsins……119
Fig. 4.6: Natural to Narrow light environment spectra and quantum catch……...120
Fig. 4.7: Narrow to Broad light environment spectra and quantum catch……121
Fig. 4.8: A summary of current knowledge about Lake Malawi *SWS1* opsin expression in the context of development and plasticity……....122
Fig. 5.1: Fine-mapping of the QTL on LG5……..136
Fig. 5.2: Quantitative PCR of the three best candidate genes on LG5…….....140
Fig. 5.3: Insertion in the first intron of the *MITFA* gene…….....143
Fig. 5.4: Both *MITFA* and *RX1* work together to modulate the expression of *SWS2A*…..145
Fig. 5.5: Two-locus interaction for *SWS2A* expression in Lake Malawi species panel….146

Fig. 6.1: Gene regulatory network for *SWS1* expression…….....158

Chapter1: Introduction

1.1 INTRODUCTION

Since the publication of Darwin's groundbreaking work 'On the Origin of Species' nearly 160 years ago, biologists have sought to understand the origin and maintenance of the vast phenotypic diversity around us [1]. The Modern Synthesis Movement coupled with the explosion in the field of molecular biology in the mid 1900s, firmly established the genetic basis of phenotypic evolution. Genetic mutation was seen as the source of new selectable variation and the elucidation of the mechanisms by which a particular genotype resulted in a given phenotype became the main question in evolutionary biology [2]. However, considering mutation to be the primary driver of variation led to a bias that ignored the reality of evolution- that natural selection acted on phenotypes and not genotypes [3]. And because the environment that an organism exists in plays a major role in modulating its phenotype, the environment also plays a substantial role in generating phenotypic variation [4, 5]. Phenotypic plasticity, defined as the capacity of a single genotype to produce multiple phenotypes in response to environmental variation, once thought to be "Lamarckian-like" [6] thus became included in what was till then a "gene"-centric evolutionary theory [2, 6-8] (Fig. 1.1).

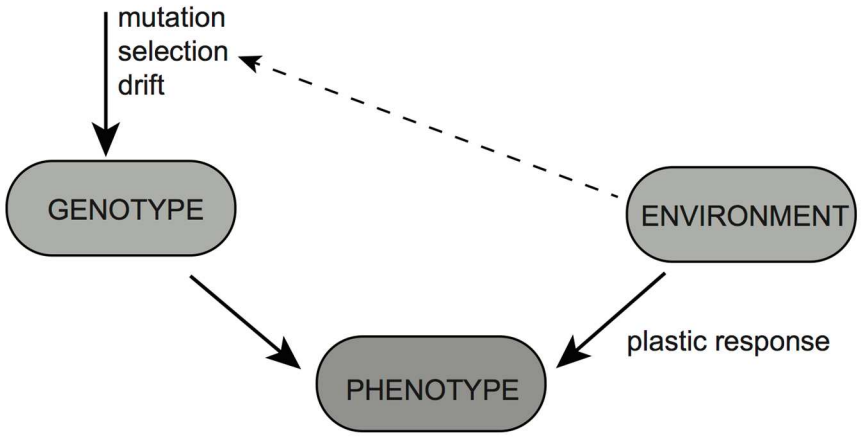


Fig. 1.1: Genetic and environmental inputs shape phenotypes. Genetic changes can lead to phenotypic differences. Under altered conditions, the environment can also induce the production of new phenotypes, which selection can then act upon.

Today, even a superficial glance at the appropriate literature will indicate hundreds of studies that show gene-by-environment interactions in a variety of organisms [9-11]. A well-known example is the production of melanin pigment. In vertebrates, melanin is produced at the end of a series of complex biochemical reactions in the melanocytes. Extensive studies in mice have delineated the genes that play distinct roles at each step of this complex pathway [12]. Mutations in any of these genes lead to the loss of the pigment phenotype to varying degrees. However, melanin pathway has also been shown to be influenced by environmental factors [13]. For example, colder temperatures tend to increase melanin production and deposition to benefit heat retention. Another example is offspring size in Trinidadian guppies. Guppies from low predation headstream sites have fewer but significantly larger offspring than counterparts from high predation downstream locations. Though these differences have been shown to have a strong genetic basis, there exists considerable plasticity in offspring size [14, 15]. Female guppies from either location when raised on re duced food rations often have offspring that are 20% larger than females raised on a constant supply of food [14]. Similarly, lacustrine species of stickleback show differences in trophic morphology, specializing either in zooplankton or zoobenthos. While sympatric species pairs show clear genetic differences, foraging environment to which the species are exposed to also has a direct effect on the morphology of the mouth parts [16]. While most of these studies have focused on morphological traits, sensory systems, such as vision respond to both genetic and environmental control as well.

1.2 VISUAL SYSTEMS

Visual communication in animals proceeds by the generation of a visual signal that gets transmitted through the environment and then detected by the receiver [17]. Thus, the environment acts as a channel within which visual systems are optimized for efficient signal perception. To adapt to environments in constant flux, animals may require plasticity in visual sensitivity to rapidly adjust to the changes [7, 18]. Yet, stable environments may select for fixed genetic differences to permanently alter visual sensitivities [19]. Understanding how visual systems are tuned both by genetic alterations as well as environmental plasticity is crucial towards understanding how sensory systems evolve and how they promote speciation.

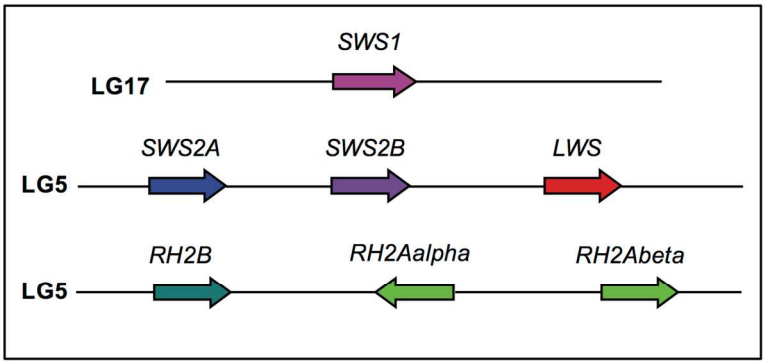
Visual response is driven by photoreceptors that are specialized neurons in the retina dedicated to transducing light. Response of a photoreceptor to photons is mediated by transmembrane proteins called opsins, together with a vitamin-A derived chromophore [20, 21]. Absorption of light leads to a conformational change in the chromophore causing the opsin to turn on a secondary signal transduction cascade resulting in neural output. Visual systems provide an excellent platform in which to study the interplay between genetic control and environmental-induced plasticity. First, the molecular basis of photoreceptor sensitivity is well-understood [22, 23]. The wavelength of light that a particular photoreceptor is most sensitive to depends on the amino acid configuration of the opsin protein it contains. Changes in a few amino acid residues, especially in the retinal binding pocket, result in drastic shifts in the sensitivity of the visual pigment [24, 25]. Second, most organisms tune their visual sensitivities to match the most abundant wavelengths of the

light environment that they exist in, a phenomenon called spectral tuning [26]. This spectral tuning can be achieved via several mechanisms such as coding sequence changes or by regulation of opsin expression in the long term but also via plasticity across shorter time scales such as due to seasonal alterations or change in water quality [24, 25, 27-34]. Last, variation in visual sensitivity, both within and between species has been shown to be conspicuously affected by natural selection. Much work has shown that visual systems are optimized for tasks such as predator avoidance and identification of food sources [35, 36]. Also, in some organisms, such as butterflies and fish, visual sensitivities form the basis for female preference in mate choice [37, 38]. Additionally, it has been established that that visual sensitivities differences can lead to population differentiation and even speciation via sensory drive [39-42].

1.3 VISUAL SYSTEMS IN AFRICAN CICHLIDS

For my thesis, I have chosen to focus on the visual systems in African Cichlids that have some of the most diverse visual systems among vertebrates on the planet. The molecular basis of visual sensitivity in African cichlids is derived from seven cone opsin genes [43]. Each of the seven cone opsin genes are maximally sensitive to a distinct part of the light spectrum and include three single cone opsins-*SWS1* (ultra-violet sensitive), *SWS2B* (violet-sensitive)*, SWS2A* (blue-sensitive), and four double cone opsins- *RH2B* (blue-green sensitive), *RH2Aβ, RH2Aα* (green-sensitive), and *LWS* (red-sensitive) [44, 45] (Fig. 1.2). As a consequence of their divergence from a riverine ancestor in the last million years, coding sequences of the opsin genes are highly conserved among species, with only a few key amino acid substitutions, especially in the opsins on either end of the spectrum- *SWS1*

and *LWS* [46-48]. Site-directed mutagenesis studies have revealed that these changes are relatively minor, causing the sensitivity of the opsin to shift by just a few nanometers (<10nm). Large differences in visual sensitivities among closely-related species in cichlids is achieved not by coding sequence changes but by altering the complement of expressed cone opsin genes [49, 50] (Fig. 1.3). Adult species consistently express only a subset (usually three) of the seven genes in their retinas at any given time. This differential expression is the most dominant mechanism for spectral tuning in cichlids and has been shown to be under genetic control.



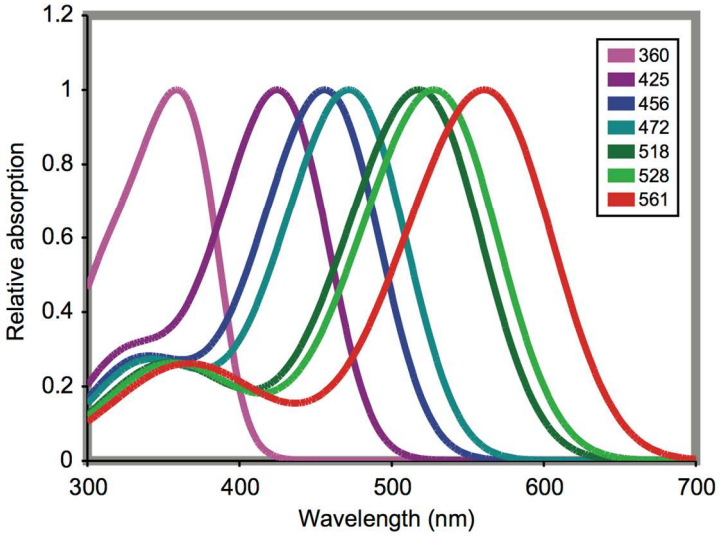


Fig. 1.2: Cichlid cone opsin genes. The seven cone opsin genes in cichlids are arranged on LG17 and two tandem arrays on LG5. Each cone opsin is maximally sensitive to a distinct part of the light spectrum, ranging from ultra-violet to red. The sensitivities of each opsin are provided in the inset.

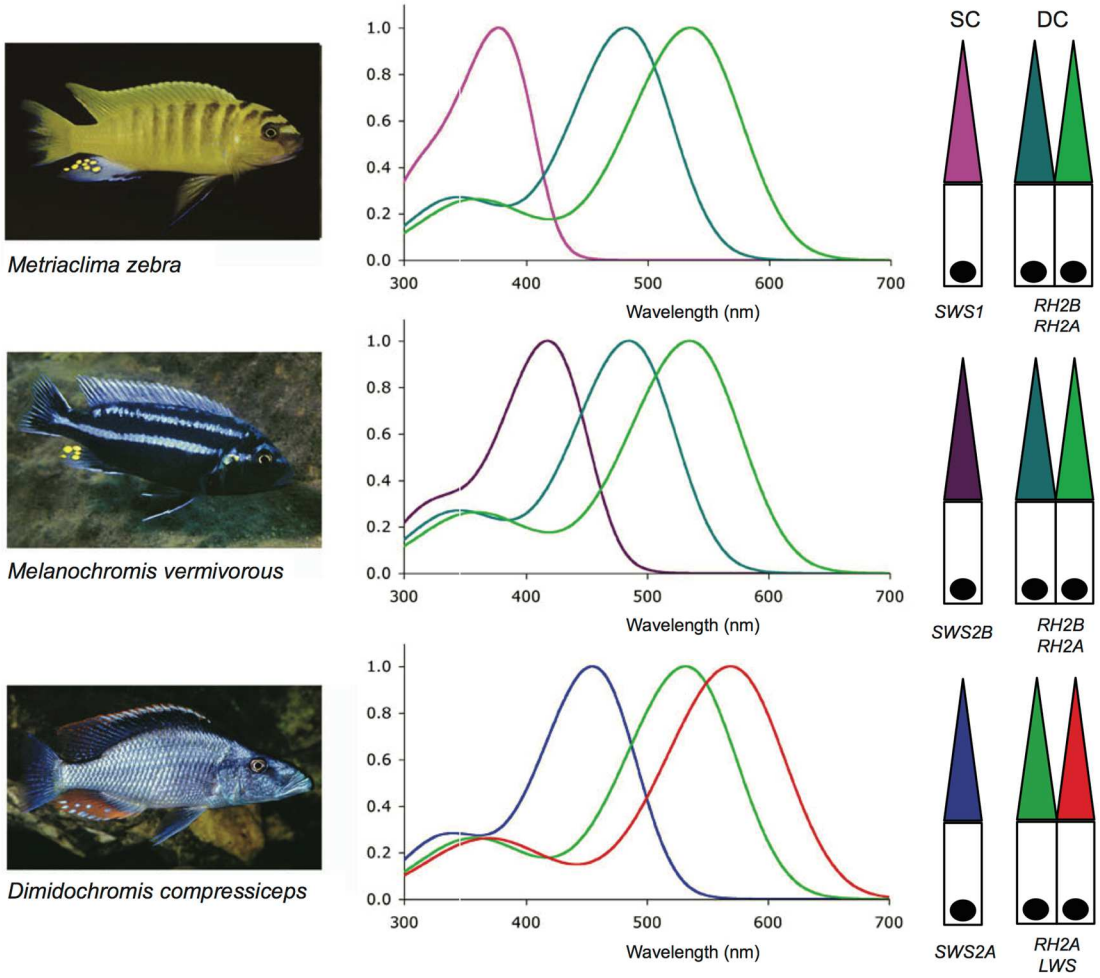


Fig. 1.3: Visual pigment templates in three Lake Malawi species. Closely-related species from Lake Malawi often have dissimilar cone opsin expression profiles. The spectral sensitivities of the visual pigments for each of the species is shown. Also depicted is opsin distribution in the single cones (SC) and double cones (DC) for each species.

Opsin expression in cichlids also changes appreciably through development. For example, single cones in ancestral Tilapia (*O. niloticus*) first express *SWS1* opsin, then progress to *SWS2B* and lastly switch to *SWS2A* opsin expression by six months, when the fish reach sexual maturity [51]. In other words, the single cones show a developmental progression

from ultraviolet to violet and finally blue sensitivity by adulthood. It is hypothesized that this progression might be a reflection on the changes in feeding behavior as the fish grow. Other cichlid species show comparable ontogenic shifts, and differences in adult expression are thought to be a result of heterochronic changes to the basic developmental program [52]. For instance, in *Melanochromis auratus,* the program is cut short and the photoreceptors retain *SWS2B* opsin as adults without ever proceeding to *SWS2A* expression.

In addition to the genetically-determined expression differences, cichlids tune their spectral sensitivities in response to changes in environmental light conditions (Fig. 1.4). Plastic changes in opsin expression have been observed in Lake Malawi cichlids during development. When reared under fluorescent light conditions in the lab, some species show small but significant alterations in opsin expression when compared to their wild counterparts [53]. Further, behaviorally-relevant changes in opsin expression were observed when individual broods were split and raised under different lighting conditions [54]. Also, developmental plasticity was observed in some cichlids when the direction of illumination was modified [55]. Dorsal retinas of fish showed increased *LWS* opsin expression when raised under upwelling fluorescent light.

1.4 IMPORTANCE OF VISION TO CICHLIDS

To sum up, both genetically determined and environmentally induced variation in cone opsin expression leads to cichlid species with dramatic differences in the visual perception ability of their photoreceptors. This substantial variation in photoreceptor sensitivity is

linked to several important ecological traits [56]. First, subtle differences in cone expression are tied to the light environment. In Lake Malawi, differences in opsin expression are often associated with depth [57]. For example, concurrent with decreasing levels of red light with increasing depths in the water column, wild *Copadichromis* species sampled from the top 5 meters of the lake have higher *LWS* expression than when sampled from 20 meters. Second, opsin expression differences are correlated with foraging needs. In Lake Malawi, zooplanktivorous and algal feeders have higher *SWS1* expression than benthic or piscivorous feeders [58]. Sensitivity in ultraviolet range is thought to enhance contrast detection of the UV-absorbing zooplankton. Third, opsin expression is linked to behaviors associated with foraging. For instance, tests have demonstrated that the duration of prey capture in UV-sensitive cichlids is lengthened in the absence of ultraviolet light input [59]. Also, optomotor response is enhanced in cichlids with higher *LWS* expression than when compared to those with low expression of *LWS* [54].

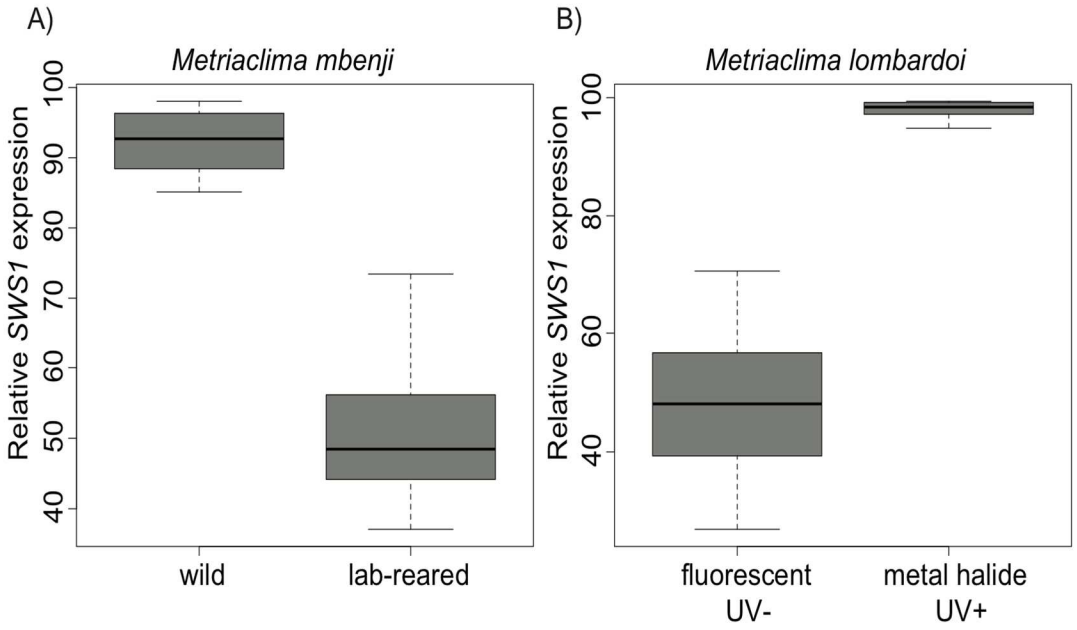


Fig. 1.4: Developmental plasticity in *SWS1* opsin expression. Lake Malawi species show changes in *SWS1* expression due to changes in rearing light conditions. A) Wild-caught *M. mbenji* have significantly higher expression of *SWS1* than lab-reared counterparts. B) Broods of *M. lombardoi* split between UV+ and UV- light environments show changes in *SWS1* expression.

Vision also plays a vital role in cichlid mating with visual signals mediating male-male aggression and female mate choice [60, 61]. In most species, males are brightly colored while the females are drab and brown. During mating season, colorful dominant males get selected by choosy females to spawn the next generation. In fact, behavioral assays testing female mate choice under altered light environments have demonstrated that while female cichlids chose the correct conspecific male under normal light, assortative mating patterns broke down when male color patterns were obscured under monochromatic light [62]. Sister taxa often differ in visual sensitivities but not morphology or behavior. Thus, it is safe to say that shifts in photoreceptor sensitivities due to opsin gene expression differences may cause phenotypic divergence in this group.

1.5 GOALS OF RESEARCH

Despite the prominent role opsins play in the evolution of phenotypic diversity in African cichlids, little is known about the genetic mechanisms controlling their differential expression. Taking advantage of the ability of cichlids to produce fertile hybrids between species and even genera, the goal is to identify the proximate genetic causes underlying differential opsin expression in this group. Previously, an intergeneric cross between two species (*Tramitochromis intermedius* and *Aulonocara baenschi*, TA cross) with varying opsin expression profiles was made in the lab [50, 63]. Quantitative genetic analysis indicated that the opsin expression differences were controlled by a few genes with large

effect [50]. Subsequently, a quantitative trait loci study conducted on F2 progeny of the cross revealed 11 loci contributing to the divergent expression of 5 cone opsins [63] (Fig. 1.5). All the identified loci were in locations far from the opsins they regulate, indicating that opsin divergence in cichlids was primarily mediated by *trans* genetic changes. Subsequent fine-mapping of one of the QTLs has shown that *Retinal anterior neural fold homeobox-1 (RX1)*, a transcription factor known to be involved in photoreceptor specification, controls *SWS2A* expression [64]. We recently identified a ~400 bp deletion in the promoter of the *RX1* gene (2.5 kb upstream of the start site) as the causative mutation (Fig. 1.5). Specifically, this deletion explains 62% of variation in *SWS2A* expression among 53 Lake Malawi species tested.

While these previous studies have, without any doubt contributed to our understanding of opsin expression divergence in African cichlids, neither parents of the earlier cross expressed *SWS1* opsin. Therefore, knowledge about the genetic mechanisms underlying the regulation of *SWS1* expression is still unknown. Expression of *SWS1* opsin in cichlids is adaptive and linked to foraging and predatory behavior [58, 59]. Finding the genetic mechanisms underlying *SWS1* differential expression thus might be the first step towards identifying the evolutionary origins of the trait.

Additionally, earlier studies have shown that cichlids are susceptible to changes in environmental light conditions during development [54, 55, 65]. While those studies have established that plasticity in opsin expression is an important tuning mechanism in these fishes, we currently do not know if adult fish can also show plasticity in opsin expression.

Plasticity is important for adult cichlids as they encounter light fluctuations in the environment due to seasonal changes in length of day or change in the clarity of water from agricultural runoff. Lack of plasticity to these environmental alterations might be detrimental to their survival. Are adult cichlids plastic and if so how is plasticity of expression translated into patterns of change in the retina and how long does it take to see those effects?

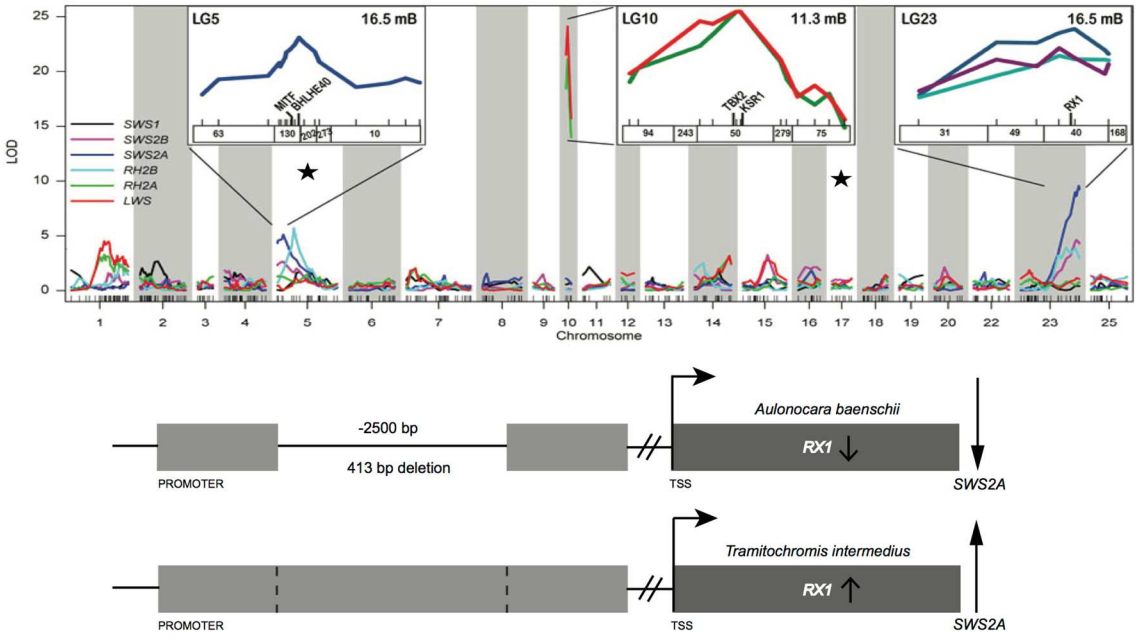


Fig. 1.5: QTL study for cone opsin expression in the *Tramitochromis* X *Aulonocara* cross. Most QTL are located in *trans* to the opsins. Insets show the significant regions for the *SWS2A*, *SWS2B*, *RH2A*, *RH2B* and *LWS* opsins. The x-axis depicts the relative position of each QTL to different linkage groups in Tilapia. Potential candidate genes within each QTL are also marked. The location of the opsins is marked with a star. (Figure courtesy: Kelly O'Quin). Deletion in the promoter of *RX1* gene is associated with a decrease in the expression of *SWS2A* opsin. While *Aulonocara* has the deletion, the mutation is absent in *Tramitochromis*.

To better understand the evolution of visual diversity in African cichlids, especially sensitivity to ultraviolet light through expression changes in *SWS1*, I have studied both genetic and environmental effects on its expression. In Chapter 2, I discuss the number of

genes controlling *SWS1* expression in a new cross between two Lake Malawi species *Aulonocara baenschi* and *Metriaclima mbenji* (MA cross) [66]. The feasibility of a QTL study to map the loci contributing to the divergent expression of opsins between the two species is also discussed. In Chapter 3, using reduced-genome sequencing, I perform a QTL analysis on the F2 progeny of the MA cross to find the loci segregating for the differential expression of opsins, with a particular emphasis on *SWS1* (Nandamuri et al., 2018, In review). In Chapter 4, I examine plasticity in *SWS1* expression due to changes in environmental light conditions [67]. Comparing two complementary experiments, I test to see if adult cichlids from Lake Malawi adjust to changes in ultraviolet illumination by altering their opsin expression. In Chapter 5, I fine-map the second QTL for *SWS2A* expression in the TA cross to determine the best candidate controlling its differential expression. Further, I highlight the interaction between one of the likely candidates and *RX1* in mediating *SWS2A* expression together. Lastly, in Chapter 6, I synthesize the main results obtained from this body of work and provide avenues for future research. In addition to expanding our understanding of the molecular landscape of opsin expression, results from this work will shed light on the evolution of visual diversity in African cichlids and how it contributes to their rapid adaptive divergence.

Chapter 2: Determination of the genetic architecture underlying short wavelength sensitivity in Lake Malawi cichlids.

Previously published under: Nandamuri S, Dalton B, Carleton K: Determination of the genetic architecture underlying short wavelength sensitivity in Lake Malawi cichlids. *Journal of Heredity* 2017, 108(4):379-390.

2.1 ABSTRACT

African cichlids are an exemplary system to study organismal diversity and rapid speciation. Species differ in external morphology including jaw shape and body coloration, but also differ in sensory systems including vision. All cichlids have seven cone opsin genes with species differing broadly in which opsins are expressed. The differential opsin expression results in closely related species with substantial differences in spectral sensitivity of their photoreceptors. In this work, we take a first step in determining the genetic basis of opsin expression in cichlids. Using a second- generation cross between two species with different opsin expression patterns, we make a conservative estimate that short wavelength opsin expression is regulated by a few loci. Genetic mapping in 96 F2 hybrids provides clear evidence of a *cis*-regulatory region for *SWS1* opsin that explains 34% of the variation in expression between the two species. Additionally, in-situ hybridization has shown that *SWS1* and *SWS2B* opsins are coexpressed in individual single cones in the retinas of F2 progeny. Results from this work will contribute to a better understanding of the genetic architecture underlying opsin expression. This knowledge will help answer long-standing questions about the evolutionary processes fundamental to opsin expression variation and how this contributes to adaptive cichlid divergence.

2.2 BACKGROUND

Understanding the genetic basis of adaptive phenotypic change has long been a central theme in evolutionary biology. Most traits of ecological relevance are quantitative and likely have a complex genetic basis. Despite the complex genetic architecture, significant progress has been made in elucidating the underlying genes, and in a few of the genetically simpler cases, causative mutations responsible for phenotypic evolution have been identified. Notable examples include male wing pigmentation patterns in *Drosophila*, skeletal reduction in sticklebacks, branching structure in maize and coat color in beach mice, among others [68-71]. Unraveling the genes responsible for phenotypic variation has provided evolutionary biologists a wealth of information regarding multiple aspects of evolution. First, the molecular nature of the identified mutations has helped delineate the connection between genotype and phenotype [68] [71]. Second, the relative contribution of the different kinds of mutations (coding changes or *cis/trans* regulatory mutations) to phenotypic divergence has been revealed [68-71]. Third, the number, distribution and the effect sizes of the contributing loci have been identified [69, 71]. Fourth, the elucidation of causative mutations has helped draw evolutionary inferences regarding the origin of the phenotypic trait [68, 69, 71]. Although causative loci of large effect have been identified, many of these involve morphological change and may not be characteristic of a broader set of traits [72, 73]. We have chosen to focus on sensory systems, as they may be somewhat complimentary to morphology as well as key to organismal behavior.

Sensory systems such as vision are highly amenable for genetic dissection because of the ease with which phenotype-genotype relationships can be tied to the ecology of the animal.

Opsins, the workhorses of the visual system, are transmembrane proteins that together with a covalently coupled chromophore facilitate the perception of light [20]. Absorption of light leads to a conformational change in the chromophore, initiating a signal transduction cascade resulting in neural output. The amino acid composition of the opsin, particularly those that interact with the chromophore in the retinal-binding pocket, determine its visual sensitivity (λ_{max}-the wavelength at which the visual pigment is most sensitive) [21]. Many studies have pointed out that the sensitivities of visual pigments in the photoreceptors closely match their photic environment and are under selective pressures to enhance visual performance in different light conditions [30, 74, 75]. Also, much work has been done to show that visual systems are optimized for tasks such as predator avoidance and identification of food sources [36]. Additionally, in some organisms like fishes and butterflies, it has been established that visual pigments form the basis for female preferences in mate choice, thereby potentially contributing to phenotypic divergence [37, 38].

African cichlids are a classical example of adaptive radiation and provide an ideal template for studying the genetic basis of evolutionary change [76-78]. These teleost fish are found throughout the rivers and lakes of Africa but are mostly focused in the three East African Great Lakes, namely Tanganyika, Malawi and Victoria [79]. Each of these lakes display incredible variety, together exhibiting more than 1000 species that have arisen in the last few million years, making cichlids the largest extant vertebrate radiation [80-82]. Cichlid species from the Great Lakes display unparalleled diversity in many morphological, behavioral and physiological phenotypes such as body coloration, trophic morphology and

parental care, among others [83, 84]. One of the most impressive of these traits includes variation in their visual sensitivities that stems from the differential expression of cone opsin genes (the genes responsible for color vision) [45, 46, 85].

All cichlid species have seven cone opsins, each conferring maximal sensitivity to a distinct part of the light spectrum [43]. These seven cone opsins include *SWS1* (ultraviolet-sensitive), *SWS2B* (violet-sensitive), *SWS2A* (blue-sensitive), *RH2B* (blue-green-sensitive), *RH2Aα* and *RH2Aβ* (green-sensitive), and *LWS* (red-sensitive). Similar to other fish, cichlids have both single and double cones in their retinas. *SWS1*, *SWS2B*, and *SWS2A* opsins are expressed in the single cones; the expression of *RH2B*, *RH2A* and *LWS* opsins is confined to the double cones [44, 45, 51]. At any given time, adult cichlid species mainly express a subset (usually three) of the seven cone opsins in their retinas leading to large shifts in their spectral sensitivities. In Lake Malawi, we refer to the three main gene combinations as: short (*SWS1*, *RH2B*, *RH2A*), medium (*SWS2B*, *RH2B*, *RH2A*) and long (*SWS2A*, *RH2A*, *LWS*) [49, 50]. Additionally, some cichlid species show plastic changes in opsin expression induced by differences in light environments [53]. Wild-caught Lake Malawi cichlids, when raised under fluorescent lights in the lab displayed small but significant changes in opsin expression, suggesting that rearing conditions may alter overall visual sensitivity.

To sum up, heterogeneity in cone opsin expression (both genetically determined and environmentally induced) may have led to the evolution of closely related cichlid species that have their photoreceptors tuned to different parts of the light spectrum. The substantial

variation in photoreceptor sensitivity is intimately connected to species ecology and serves to fine-tune cichlid vision in response to environmental and behavioral factors such as ambient light, foraging and habitat preference [56, 59]. Vision plays a vital role in cichlid mating with visual signals mediating male-male aggression and female mate choice [60, 61]. For example, behavioral assays have demonstrated that while female cichlids exhibited species specific assortative mate choice when male color patterns were visible under normal light, non-assortative mating was observed when conspecific male pigmentation patterns were obscured under monochromatic light [62]. Furthermore, plasticity in opsin expression may have improved the ability of cichlids to rapidly colonize new environments and facilitate ecological diversification [53]. As such, any shifts in photoreceptor sensitivities due to opsin gene expression differences may contribute significantly to phenotypic divergence in this group.

Significant progress has been made to identify key transcription factors that are critical for photoreceptor differentiation. Much of this information has come from work using mouse knock-outs to test the effects of specific genes, but important factors have also been identified in other systems such as chicken and *Xenopus* [86, 87]. For example, transcription factors such as retinoic acid receptor (*RAR*) are known to influence opsin gene expression in mice, fruit flies, fish and humans [88-91]. Fish typically have more cone opsins with representatives from all four vertebrate classes (*SWS1*, *SWS2*, *RH2*, *LWS*; cichlids have seven cone opsins, and zebrafish have eight [92]. Although it is well known that the RH2 and LWS tandem opsin arrays in zebrafish are mediated by *cis*-enhancers, work to understand the molecular mechanisms underlying the differential expression of

cone opsin genes is just beginning [93-95]. The existence of closely related species with variable opsin expression and the ability to make fertile hybrids makes cichlids a powerful system in which to study the genetic basis of opsin expression divergence in naturally occurring populations.

Despite the prominent role opsins play in the evolution of phenotypic diversity in African cichlids, little is known about the genetic mechanisms directing their expression. Due to their recent evolutionary history, all species from Lake Malawi have nearly identical coding sequences for all the seven cone opsins [46]. Preliminary studies of the opsin proximal promoters and non-coding elements located close to opsins suggest that these regions are highly conserved [96]. Large differences in visual sensitivities in cichlids are achieved not by changing the coding sequence of the opsin but by altering the combination of expressed opsin genes [49, 50]. To identify the genes responsible for the differential expression of cone opsins, a genetic cross between two species (*Tramitochromis intermedius* and *Aulonocara baenschi*) with varying opsin combinations was previously made in our lab [50, 63]. A quantitative trait locus study conducted on the F2 progeny of the cross revealed five loci contributing to the opsin expression differences between the two parental species [63]. All five of these identified loci are located in *trans* far from the opsin genes. Subsequent fine-mapping of one of the loci has indicated that the retinal anterior neural fold homeobox-1 (*Rx1*), a transcription factor known to be involved in photoreceptor specification, controls *SWS2A* expression [64]. We recently identified a ~400bp deletion in the promoter of the *Rx1* gene (2.5kb upstream of the start site) as the

causative mutation. Specifically, this deletion explains 62% of variation in *SWS2A* expression among 53 Lake Malawi species (with different opsin expression patterns) [64].

In the present study, we have made another experimental cross between two Lake Malawi species - *Metriaclima mbenji* and *Aulonocara baenschi* (MA cross). The goals of this study are four-fold- 1) We aim to characterize opsin expression variation in the parents (F0) as well as the F1and F2 of the cross. 2) We want to estimate the number of genes regulating opsin expression differences between the two species. This will help with establishing the feasibility of a future quantitative trait locus (QTL) study to map contributing genes. Due to the ease with which opsin expression changes from one species to another within the recently arisen Malawi flock, we hypothesize that opsins are controlled by a few factors of large effect. 3) We test for association with genetic markers located in close proximity to the opsins to check for evidence of *cis*-regulation to explain the differences in opsin expression. 4) We utilize in-situ hybridization to show that single cone opsins are coexpressed in individual cone cells in a sample of F2 progeny. Results from this study provide insights into the molecular mechanisms by which opsin expression varies throughout the cichlid flock of Lake Malawi.

2.3 MATERIALS AND METHODS

2.3.1 Parental Species

Metriaclima mbenji (MM) *and Aulonocara baenschi* (AB) are two species from Lake Malawi that diverged from a common ancestor less than a million years ago [97]. In the lake, both species occupy different microhabitats and have different styles of feeding.

While MM are omnivorous rock-dwellers, AB live in habitats intermediate between sand and rocks where they forage predominantly on small invertebrates [98]. Despite their very different habitat and behavior, *Aulonocara* are phylogenetically close to *Metriaclima* [99]. In terms of opsin expression, wild-caught MM express the short gene combination with *SWS1* in the single cones plus the *RH2B* and *RH2A* in the double cones. When previously reared in the lab under fluorescent lights, substantial shifts in expression have been observed in both single and double cones [53]. Wild-caught AB have the medium opsin combination with *SWS2B* in the single cones and expression of *RH2B* and *RH2A* in the double cones. Whether AB show changes in opsin expression when raised in the lab is currently unknown and needs to be determined.

Retinas from wild-caught individuals used in this study were collected during a field expedition to southern Lake Malawi. MM individuals (n=10) were sampled from Mbenji Island (13º 26'23.50"S 34º29'35"E) in 2008 and were used previously in Hofmann et al., 2010 [53]. AB individuals (n=7) were obtained from Nkhomo reef (13º0.964'S 34º20.114'E) in 2012. Immediately after capture, the fish were euthanized and retinas were dissected at the field research station and stored in RNAlater for subsequent phenotyping. During the 2008 field trip, live adult MM collected from the same population were brought back to our aquaculture facility at the University of Maryland and bred in the lab for one generation. Lab-reared AB were obtained from the aquarium trade in 2005 and bred in the lab for a few generations. Subsequently, adult lab-raised MM (n=10) and AB (n=10) sampled from individual broods of each species were phenotyped for opsin expression [50, 53]. All sampling for both wild-caught and lab-reared individuals was conducted in random

with respect to sex. Student t-tests were performed in R version 3.2.3 (R Foundation for Statistical Computing, Vienna, Austria) to test for significant differences ($P< 0.05$) in the mean expression of opsins between wild-caught and lab-reared fish.

2.3.2 Hybrid cross

Cichlid species will often interbreed in lab conditions to produce fertile progeny. F1 hybrids of the MA cross were obtained by crossing a single MM female to a single AB male. MM parental specimen used for the cross was third generation lab-raised progeny generated from a wild-caught stock maintained in the aquaculture facility. The AB parent of the cross was multiple generation progeny generated from fish obtained from the aquarium trade. Around 40 F2 families (~350 F2 hybrids) were spawned from 11 F1 parents- one sire and ten dams breeding freely. In the present study, we used a total of 210 individuals- 10 each of the parental species (MM and AB), 24 F1 and 166 F2 hybrids comprising 12 F2 families. Sampling of F1 and F2 hybrids was random with respect to sex. All animals used in this study were at least six months old (Cichlids attain sexually maturity at six months of age and are considered adults). It should be noted that the F0 individuals used in this study could be older than the F2 hybrids. However, previous work in other closely related lab-reared cichlids (*M. zebra*, *D. compressiceps*, *T. intermedius* and outgroup Tilapia) showed that opsin expression is labile during development, but remains unchanged once they attain adulthood [51].

2.3.3 Phenotyping opsin expression and correlations among opsin genes

After euthanizing the animals with MS-222, whole retinas were isolated from both eyes and placed in RNAlater. Cichlid opsin expression has been shown to vary with time of day, though the opsins are generally correlated in their circadian variation [100]. To minimize variation in opsin expression due to time of day, all dissections were performed between 3PM and 5PM. Total RNA was extracted from the dissected retinas using the QIAshredder and RNeasy kits (Qiagen, Valencia, CA, USA). 0.5µg of the total RNA was reverse transcribed with Superscript III (Invitrogen, Carlsbad, CA, USA) to make cDNA. This cDNA template was then used to quantify gene expression with quantitative RT-PCR utilizing opsin-specific Taqman primers and probes according to our previously published methods [43, 50]. RT-PCR Taqman primers and probes for each opsin are provided in Table 2.1. Similar to our earlier studies, a construct containing six opsin fragments was used as a control. All RT-PCR reactions were performed in duplicate. Single and double cones differ in which opsins they express; therefore, gene expression was normalized as a percent of total opsin expression within each cell type separately. For example, the fraction, *f*, of *SWS1* expression in the single cones is calculated as:

$$f_{SWS1,\,SC} = T_{SWS1} / (T_{SWS1} + T_{SWS2B} + T_{SWS2A})$$

where T is the transcript level for each of the single cone opsin genes. Similarly, the fraction of LWS opsin expression in double cones is:

$$f_{LWS,\,DC} = T_{LWS} / (T_{RH2A} + T_{RH2B} + T_{LWS})$$

Since the expression of each opsin was normalized to the total opsin expression in single and double cones separately, we expect to find some inherent correlations in expression between the genes of each cone type. Regardless, correlation analyses among F2 expression of all opsins were performed in R (R Foundation for Statistical Computing,

Vienna, Austria). Correlations were deemed significant for $P < 0.05$ and also corrected for multiple comparisons (Bonferroni correction).

2.3.4 Calculating number of genes controlling expression

The Castle-Wright estimator was used to calculate the number of genetic factors mediating opsin expression [101, 102]. According to this method, the number of loci responsible for phenotypic differences between the parental lines is:

$$N_e = [(\mu_{MM} - \mu_{AB})^2 - \sigma^2_{MM} - \sigma^2_{AB}]/8(\sigma^2_{F2} - \sigma^2_{F1})$$

where N_e is the number of effective factors, μ is the mean gene expression in each parental species MM or AB, and σ is the variance in gene expression in the parentals (MM and AB), F2 and F1. As the opsins are normalized relative to each cone type, the Castle-Wright estimates obtained for each opsin within single and double cones are not independent. The estimator is based on multiple assumptions, including that the loci are unlinked, all loci have equal allelic effects, the alleles act additively and that the alleles increase the trait in one parent and decrease it in the other. It is possible that one or more of these assumptions might be violated in the cross leading to an underestimation of the number of factors.

Additionally, we also estimated the degree of dominance in this cross from the mean expression in F1 hybrids and the parentals. Dominance is calculated individually for each opsin gene as:

$$d = [(\mu_{F1} - \mu_{MM})/(\mu_{AB} - \mu_{MM})] - 0.5$$

where μ is the mean expression in F1 or in one of the parents, AB and MM. In the presence of dominance, the effective number of factors as calculated by the Castle-Wright estimator changes. If h=d+0.5, the number of factors now become:

$$N_{e,\,dom} = [1.5\text{-}2h(1\text{-}h)]\, N_e$$

2.3.5 Localization of opsin in the single cones

To determine whether individual singe cone cells coexpressed *SWS1* and *SWS2B*, we performed dual-labeling fluorescent in-situ hybridization on whole retinas of four F2 hybrid individuals according to published methods [103, 104]. Briefly, following euthanasia, retinas were removed and fixed in 4% paraformaldehyde. We used DIG or fluorescein labeled riboprobes that were designed specifically to the full coding sequences of *SWS1* or *SWS2B*. Probe signals were localized by sequential tyramide signal amplification using Alexa Fluor 594 and 488 dyes (Invitrogen, Carlsbad, CA, USA) and visualized on a Leica DM 5500B epifluroescence microscope (Leica Microsystems, Wetzlar, Germany). In each individual, 50 single cones were identified by their morphology and examined to determine whether they expressed either or both *SWS1* and *SWS2B*.

2.3.6 Testing for cis-regulation

To check for evidence of *cis*-regulation contributing to the differences in opsin expression between the two species, we genotyped microsatellite markers located in close proximity to the opsins. In cichlids, the opsin genes are positioned in three genomic regions [105]. The *SWS1* gene is located on linkage group (LG) 17, *SWS2A*-*SWS2B*-*LWS* and *RH2B*-

RH2Aα-RH2Aβ form two separate tandem arrays on LG5. The *SWS2A* tandem array is 18kb long and separated from the 26 kb long *RH2B* array by 30cM. The details of each marker (location and primer sequences) are provided in Table 2.2. The cichlid genetic map is 1300cM long with each cM roughly corresponding to about 750kb (genome size~10^9 bp) [105]. Thus, all the microsatellite markers used in this study are well within 1cM and therefore are in tight linkage with the opsins. Around 96 individuals were genotyped for each microsatellite marker and association between marker genotype and opsin expression was tested using analysis of variance (ANOVA) with a significance level of $P < 0.05$ in R (R Foundation for Statistical Computing, Vienna, Austria). For any positively associated marker, a regression analysis of the phenotype (opsin expression) with associated marker genotype was subsequently performed in R to quantify the fraction of variation in expression.

2.4 RESULTS

2.4.1 Opsin expression in the parental species

Consistent with the previous study [53], qRT-PCR results indicate expression differences between lab-reared and wild-caught MM individuals (Fig. 2.1A). Noteworthy changes are observed along the *SWS1-SWS2B* axis, with a significant decrease in the expression of *SWS1* (t-test, $P= 2.01E^{-9}$) and an increase in *SWS2B* expression (t-test, $P= 1.55E^{-10}$) in lab-reared fish. Changes are also observed along the *RH2B-RH2A* axis in the double cones with an increase in *RH2B* (t-test, P= 0.00009) and a decrease in *RH2A* expression (t-test, $P= 7.82E^{-6}$). Expression of *LWS* opsin increased slightly in the lab MM, but was not statistically significant. On the other hand, no significant differences in opsin expression

were observed between wild and lab raised AB individuals (Fig. 2.1B). Only in *SWS2A* were there statistically significant differences between lab-reared and wild-caught individuals, but the levels of *SWS2A* expression in both AB and MM are likely too low (less than 1%) to contribute to visual function.

Table 2.1: Opsin-specific Taqman primer and probes used for RT-PCR. For each opsin, the forward primer, reverse primer and the probe are provided.

Opsin	Oligo name	Sequence
SWS1	CicUVF.tq	GGCTGTGCCTGCCCAC
	CicUVR.tq	AGGAGCAGCCCAGACCTTC
	CicUV.tq	TTTCTTTGGCTGGAGCAGGTACATCCC
SWS2A	CicB1F.tq	TTTGGTGCGCTAGCATGC
	CicB1R.tq	AAGGGACCACAGGCTTACCAT
	CicB1.tq	AGATCGAAGGTTTCATGGTAACACTCGGTG
SWS2B	CicB2F.tq	GCGCTGCACTTCCACCTC
	CicB2R.tq	GGCCACAGGAACACTGCAT
	CicB2.tq	TTGGATGGAGCAGGTATATCCCAGAGGG
RH2A	CicGrnF.tq	TTCTGTGCAATTGAGGGATTC
	CicGrnR.tq	CCAGGACAACAAGTGACCAGAG
	CicGrn.tq	TGGCCACACTAGGAGGTGAAGTTGC
RH2B	CicGreen3F.tq	TGCTGCCCCCCCATTG
	CicGreen3R.tq	AGGTCCACAGGAAACCTGAA
	CicGreen3.tq	TGGCTGGTCAAGGTACATTCCTGAGGGA
LWS	CicRedF.tq	CTGTGCTACCTTGCTGTGTGG
	CicRedR.tq	GCCTTCTGGGTTGACTCTGACT
	CicRed.tq	TGGCCATCCGTGCTGTTGCC

2.4.2 Opsin expression in the hybrid cross

Opsin gene expression was highly variable among the 24 F1 and 166 F2 individuals in the MA cross. The means and variances of expression values for the single (*SWS1*, *SWS2B*, *SWS2A*) and double cone opsins (*RH2B*, *RH2A*, *LWS*) are given in Table 2.3. As expected

from the F0 (MM and AB), there was no significant expression of *SWS2A* in the F2 hybrids (Fig. 2.2A). Mean levels of *SWS1* expression in both F1 and F2 were similar to the parental AB phenotype (Fig. 2.2B). Likewise, a strong dominance was observed in the *SWS2B* expression, with both the F1 and F2 hybrids showing levels similar to that of the AB parent (Fig. 2.2C). Compared to the single cones, both F1 and F2 hybrids showed higher variance in expression levels of *RH2B*, *RH2A* and *LWS* opsins. Both the average expression of *RH2A* and *RH2B* in F2 was closer to the AB parent. (Fig. 2.2D, E). In contrast, expression of the *LWS* was transgressive, with the F2 hybrids extending slightly beyond the phenotypic range of the parents (Fig. 2.2F).

All the opsin correlations are provided in Table 2.4. As expected, in the single cones of the hybrid cross, there is a negative correlation between expression of *SWS1* versus both *SWS2B* and *SWS2A*. Similarly, in the double cones there is a negative correlation between *RH2A* versus both *RH2B* and *LWS*. In this study, there exists only one correlation between single and double cone opsins that is independent of normalization- a significant negative correlation between *RH2B* and *SWS2A* that is not likely to be biologically relevant due to the low *SWS2A* expression.

Table 2.2: Genotyping primers for markers located in close proximity to the opsin genes. The sequences and location of each marker relative to the closest opsin genes are listed.

Marker	Location	Forward primer 5'-3'	Reverse Primer 5'-3'
SWS1_37372 $(TCAA)_9$	LG17; located in the third intron of the *SWS1* gene	CACGACCAGCTACACAC ACT	TGACGAAGGACCCCAC CATA
LG5_49353 $(CA)_{34}$	LG5; located 31 kb upstream of the *SWS2A-SWS2B-LWS*	TGAAGGCGGTGGTCTTA TCT	TGGGTGTTGGTCAGAA ACAG
LG5_14393 $(AAT)_{20}$	LG5; located 28 kb upstream of the *RH2B-RH2Aβ-RH2Aα*	AGGACAGCATGACCCAA AAC	GCTCCTCTCAGTGGCCT GTA

Table 2.3: Means and variances (in parenthesis) of relative opsin expression in the F0, F1 and F2 generations of the hybrid cross.

Opsin	MM (n=10)	AB (n=10)	F1 (n=24)	F2 (n=166)
SWS1	39.422 (167.313)	0.938 (1.505)	1.725 (0.685)	3.556 (29.199)
SWS2B	60.56 (167.32)	98.287 (2.019)	98.042 (0.715)	96.111 (28.469)
SWS2A	0.018 (0)	0.774 (0.279)	0.233 (0.068)	0.333 (0.388)
RH2B	25.901 (6.407)	16.366 (3.027)	13.727 (45.1)	14.522 (32.809)
RH2A	73.037 (7.999)	82.921 (4.648)	73.151 (195.056)	76.392 (115.027)
LWS	1.062 (0.174)	0.713 (1.381)	13.122 (196.739)	9.086 (91.614)

Table 2.4: Opsin gene correlations and p-values (in parenthesis) in F2 hybrids. Opsin expression is measured relative to the total opsin expression within each cone type (single cones / double cones). Correlations significant at p-value <0.05 are bolded. Correlations significant after Bonferroni correction are starred.

	SWS1	*SWS2B*	*SWS2A*	*RH2B*	*RH2A*	*LWS*
SWS1	1					
SWS2B	**-0.993*** (0)	1				
SWS2A	**-0.166** (0.032)	0.052 (0.509)	1			
RH2B	0.046 (0.555)	-0.014 (0.857)	**-0.279*** (0.0003)	1		
RH2A	0.035 (0.6583)	-0.047 (0.5477)	0.103 (0.1883)	**-0.458*** ($5.711E^{-10}$)	1	
LWS	-0.663 (0.3957)	0.061 (0.4344)	0.052 (0.5029)	-0.09 (0.2723)	**-0.847*** (0)	1

2.4.3 Castle-Wright estimates of genetic factors controlling opsin expression

Based on the means and variances in opsin expression in the three generations (F0, F1 and F2), we have calculated the number of factors mediating each opsin (Table 2.5). The Castle-Wright equation provides an estimate of 5.75 genes regulating *SWS1* and 5.64 genes

for the *SWS2B* opsin. Consistent with the low levels of *SWS2A* expression in this cross, the estimator computed that 0.11 genes are controlling it. In the double cones, for the *RH2B*, *RH2A* and *LWS* opsins, the Castle-Wright equation provided an estimate of -0.83, -0.13 and 0 genes respectively. The low and negative estimates of genetic factors for the double cone opsins arises from the fact that the expression values for these opsins exhibit high variance in F1 and F2 hybrids relative to the mean difference between the parental species.

Opsin expression is characterized by dominance in the MA cross. The calculated dominance values (d) for each opsin are as follows: *SWS1*: 0.48, *SWS2B*: 0.494, *SWS2A*: -0.22, *RH2B*: 0.78, *RH2A*: -0.49 and *LWS*: -34.99 (Table 2.5). Dominance values usually range from -0.5 to 0.5 and the skewed value obtained for the *RH2B* is indicative of the decreased mean expression in the F1 and F2 hybrids relative to the parents. Similarly, calculated dominance for *LWS* is impacted by transgressive expression in the F1 and F2 individuals. Dominance values for the *SWS1*, *SWS2B* and *RH2B* opsins are positive and closer to AB parental expression levels, dominance in *SWS2A* and *RH2A* is negative, leaning towards the expression in MM parent. Inclusion of the dominance values has a moderate effect on the number of effective factors controlling expression for *SWS1* and *SWS2B*, raising the number to 8.39 genes for both the opsins (Table 2.5). For *SWS2A*, *RH2B* and *RH2A*, dominance has negligible effect on the number of genetic factors. The number of genetic factors for *LWS* reaches 4.17, likely an artifact from the abnormal calculated dominance.

2.4.4 Opsin coexpression in the single cones

The cichlid retina is composed of a regular mosaic comprising four pairs of double cones surrounding every single cone [106]. *In situ* hybridization indicated that whenever *SWS1* was expressed, it was coexpressed in a single cone that also expressed *SWS2B* (Fig. 2.3). Although three of the four F2 hybrids tested expressed *SWS1* opsin, the frequency of *SWS1* expression varied both overall among individuals and spatially across their individual retinas. However, all single cones in all four individuals examined showed *SWS2B* expression.

Table 2.5: Castle-Wright estimates of the number of factors (N_e) controlling the expression of each opsin. We also list the dominance (d) and $N_{e,\ dom}$ estimated with effects of dominance.

Opsin	N_e	Dominance (d)	$N_{e,dom}$
SWS1	5.75	0.48	8.39
SWS2B	5.64	0.49	8.39
SWS2A	0.11	-0.22	0.13
RH2B	-0.83	0.78*	-1.83
RH2A	-0.13	-0.49	-0.19
LWS	0	-34.99*	4.17

*Dominance values usually range from -0.5 to 0.5. The abnormal dominance values are due to decreased mean expression in the F1 and F2 hybrids relative to the parents.

2.4.5 Testing for cis-regulation of opsins

Three markers, one located in close proximity to each of the three-opsin arrays (*SWS1*, *RH2B-RH2Aα-RH2Aβ*, and *SWS2A-SWS2B-LWS*), were tested in 96 F2 hybrids for association with all the opsins. Only one marker (SWS1_37372) tested significant for association with *SWS1* ($P= 1.61e^{-09}$) and *SWS2B* ($P= 8.35e^{-09}$). This positively associated marker is located in the third intron of the *SWS1* gene and suggests that *cis*-regulation might

be important for the expression of *SWS1*. However, it is to be noted that with the ~ 0.5 cM resolution resulting from genotyping 96 F2 individuals, genomic blocks of at least 400kb will be inherited together (cichlid genetic map is 1300cM with each cM roughly corresponding to 750kb). By testing for a marker located in the third intron of the *SWS1* gene, we are testing for the presence of a *cis*-acting factor located within that 400kb block. It might be possible that other gene/s located within that interval could also be contributing to the difference in *SWS1* expression. We quantified the fraction of variation in *SWS1* expression by performing a regression analysis of *SWS1* expression on genotype at the SWS1_37372 marker in the F2 individuals (MM/MA/AA; Fig. 2.4). In the hybrid cross, this tightly linked marker explained 34% of the variation. However, it should be cautioned that this number could be an overestimate as a consequence of the Beavis effect that states that the proportion of phenotypic variances associated with identified QTL are greatly overstated with a small sample of progeny [107]. We found no evidence of *cis*-regulatory factors that explain the differences in opsin expression between the two parents of the cross, for either the *SWS2A-SWS2B-LWS* or the *RH2B-RH2Aα-RH2Aβ* tandem arrays, both of which are located on LG5.

2.5 DISCUSSION

2.5.1 Plasticity in opsin expression

Using quantitative PCR, we have shown that light-induced plasticity differences occur in some cichlid species but not in others. AB showed no changes in opsin expression between wild caught and lab reared conditions but MM demonstrated significant expression differences. These observations agree with previous results in the lab that showed that the

magnitude of plasticity varies between several Lake Malawi species [53]. When compared to solar illumination (>50000 lux; decreases exponentially with depth), fluorescent lights in the lab are two orders of magnitude lower in intensity (500 lux), and more crucially, deficient in ultraviolet (UV) wavelengths (< 400nm) [53]. The net result of this UV deficiency is that the quantum catches of single cones with *SWS1* visual pigment is reduced to negligible levels. The altered light spectrum might induce the expression of additional opsins as a plastic response to increase the amount of light collected by the photoreceptors in certain species. This might explain the reason for the simultaneous decrease in *SWS1* and an increase in *SWS2B* expression in MM. *SWS1* is never expressed in adult AB, therefore no such deficits in quantum catches ever occur, so plastic changes might not occur in this species.

2.5.2 Co-expression in the single cones

Using fluorescent double-labeling *SWS1* and *SWS2B* mRNA probes; we have shown that some single cones in the retinas of F2 hybrids coexpress *SWS1* and *SWS2B* opsins. These results add to the growing list of studies that report expression of multiple opsins in single photoreceptors [108-111]. Even more importantly, these results complement our previous work to examine coexpression in both single and double cones in the closely related *Metriaclima zebra,* where we showed that coexpression exhibits a spatial distribution in the retina that tunes the sensitivity of specific retinal regions to the different background spectra that they view in the lake [103, 108-111]. It is known that the presence of two opsins in the same cell broadens its spectral range, thereby increasing the quantum catch of the photoreceptors and serving as an adaptation for the retina to catch more light [103].

This is particularly relevant in the present study as the F2 hybrids are reared in the lab where light intensity is much lower than in the wild. One potential downside to the coexpression of opsins is a decrease of spectral opponency that reduces the ability to discriminate colors [112]. Behavioral experiments that test the ability of cichlids to discriminate wavelengths are currently in progress in the lab. It would be interesting to test if F2 hybrids with varying levels of coexpression show differences in color perception. The *SWS1* opsin is located on LG17 and the *SWS2B* opsin is present on LG5 [105]. Coexpression of these two opsins in the same cell might be mediated by a common transcription factor that binds to specific sites in the promoters of both the opsins. As such, work is currently in progress to determine if the promoters of these opsins have any shared identity.

2.5.3 Opsin expression differences are controlled by a few genes

As a first step towards understanding the genetic basis of opsin expression in this cichlid cross, we used the Castle-Wright estimator to predict the number of genes underlying the interspecific differences in expression. This widely used method provides a minimum number of genetic factors underlying a trait and is based on the amount of segregational variance occurring in a cross [102]. The Castle-Wright estimator assumes fixed differences between the parents, additive gene action, equal allelic effects and unlinked loci. Despite these stringent assumptions, several modeling and quantitative trait loci studies have suggested that the estimator functions fairly well [113-115]. When the number of factors is low, sample sizes are high and the parental populations differ by several standard deviations, the Castle-Wright estimate is close to the number estimated from full QTL

analyses [113, 116]. Within cichlids, this method has been successfully applied to a variety of phenotypes such as jaw morphology and pigmentation traits [50, 83, 84, 113, 114]. In our own work, using the Castle-Wright equation, we previously estimated that around 10-12 genes underlie opsin expression differences in a hybrid cross between *Aulonocara baenschi* and *Tramitochromis intermedius* [50]. A follow-up QTL study on the F2 hybrids of the same cross, using RAD-seq tags identified 11 QTL contributing to the differing expression of opsin genes, suggesting that the Castle-Wright estimator is relatively robust and can be applied to F2 crosses between phenotypically divergent species [63]. In all cases the number of genetic factors should be considered as a lower bound on the total number of loci controlling the trait of interest.

In the MA cross, we find that the number of genes controlling the opsins is relatively small with six genes for each of the *SWS1* and *SWS2B* opsins. Both these opsins show a dominant mode of inheritance in the MA cross which on inclusion in the Castle-Wright equation brings up the number of factors to eight for each of these opsins. Consistent with negligible levels of expression in the parents as well as the hybrids, the estimator computes zero contributing factors for the *SWS2A* opsin. The skewed estimates obtained for the double cone opsins (*RH2A*, *RH2B* and *LWS*) imply that one or more assumptions underlying the Castle-Wright estimator may not have been fulfilled in this cross suggesting that perhaps the genetic architecture underlying double cone opsins is more complicated. A more complete QTL study may provide some insights as to which of the assumptions have been violated. From this point forth, all discussion will focus on the two single cone opsins, *SWS1* and *SWS2B*.

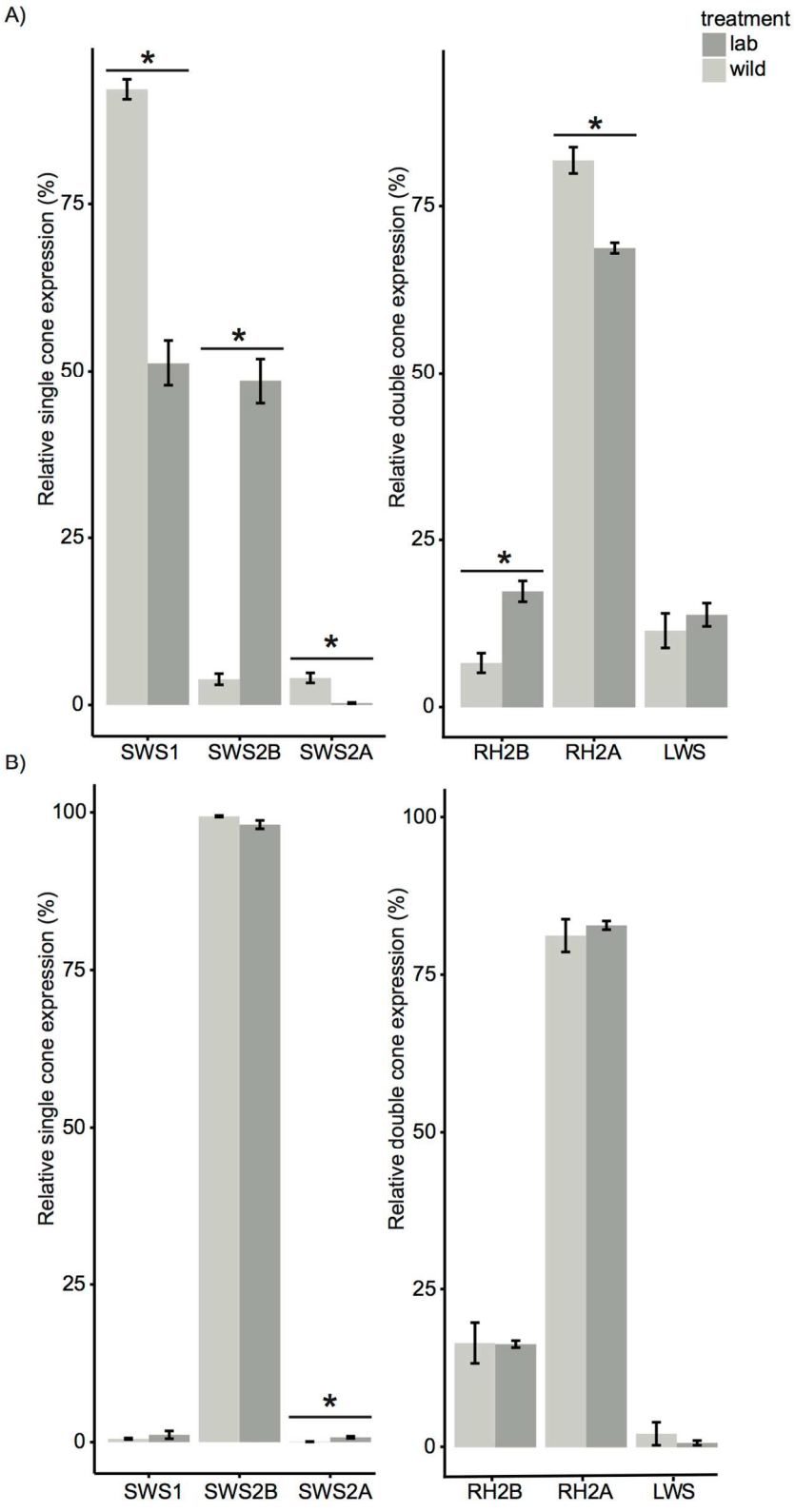

A)
treatment
lab
wild
Relative single cone expression (%)
Relative double cone expression (%)
0
25
50
75
*
SWS1
SWS2B
SWS2A
RH2B
RH2A
LWS
B)
100

Fig. 2.1: Variation in opsin expression between wild-caught and lab-reared adult cichlids. Mean relative opsin expression for *Metriaclima mbenji* (MM) (1A) and *Aulonocara baenschi* (AB) (1B) for both single and double cones are shown. Dark gray bars are lab-reared individuals and light gray are wild-caught. Opsin genes that differ significantly ($P < 0.05$) between the two treatments are marked with an asterisk. Error bars are ±SE. 1A: MM shows significant differences in expression for all opsins except *LWS*. 1B: Except for the *SWS2A* opsin, AB show no statistically significant differences in opsin expression.

Since expression of each opsin (*SWS1* or *SWS2B*) in the cross is characterized as a fraction of total expression in single cones (*SWS1*+*SWS2B*+*SWS2A*), Castle-Wright estimates obtained for the *SWS1* and *SWS2B* opsins are not independent of each other. Due to negligible levels of *SWS2A* expression in the cross, expression of *SWS1* and *SWS2B* opsins becomes inversely correlated. In other words, as expression of *SWS1* increases, *SWS2B* expression decreases and vice-versa. Thus, in the MA cross, approximately eight loci seem to be regulating the expression of the *SWS1* and *SWS2B* opsins. However, we hypothesize that the actual number of factors mediating the expression of these two opsins might be lower than the estimated number due to two main reasons. First, we show that *SWS1* and *SWS2B* are being coexpressed in the same cone in F2 hybrids. Since the two opsins are expressed in the same cell, their expression might be mediated by a common set of genetic factors. Second, the plasticity assay in *Metriaclima mbenji* individuals points to the existence of a trade-off between the expression of *SWS1* and *SWS2B* opsins. This leads us to speculate the presence of other factors that activate the expression of one opsin while repressing the other. Taken together, these reasons suggest that expression divergence along the *SWS1*-*SWS2B* axis may require coordinated changes at multiple loci, at least some of which might directly regulate the expression of both opsins.

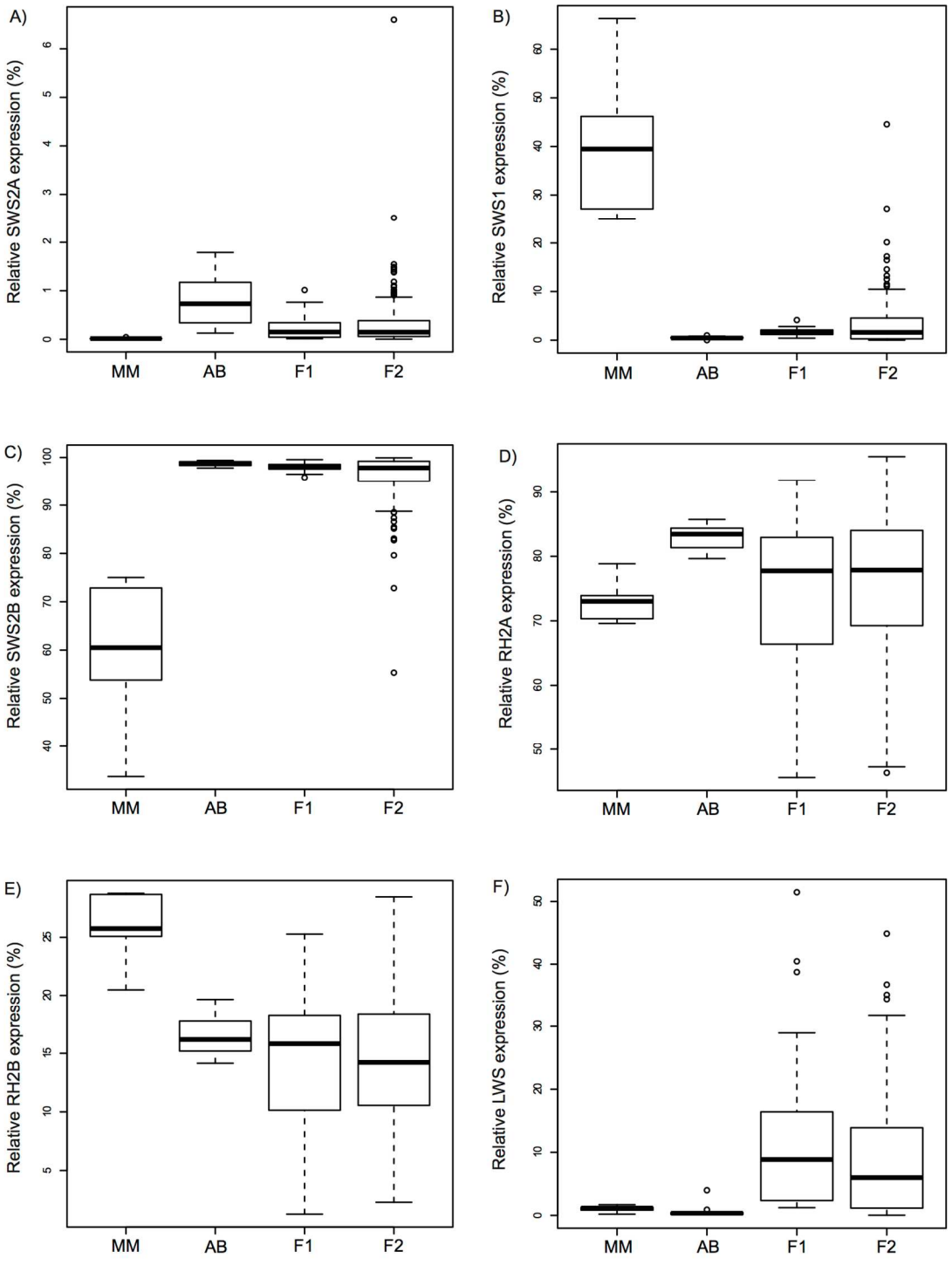


Fig. 2.2: Boxplots of opsin expression from the parental, F1 and F2 generations in the MA cross. The line within each box represents the median of the data, the areas of the box above and below the line indicate the upper and lower quartiles and the whiskers show the expression range of each opsin. A) *SWS2A* expression; B) *SWS1* expression; C) *SWS2B* expression; D) *RH2A* expression; E) *RH2B* expression; F) *LWS* expression.

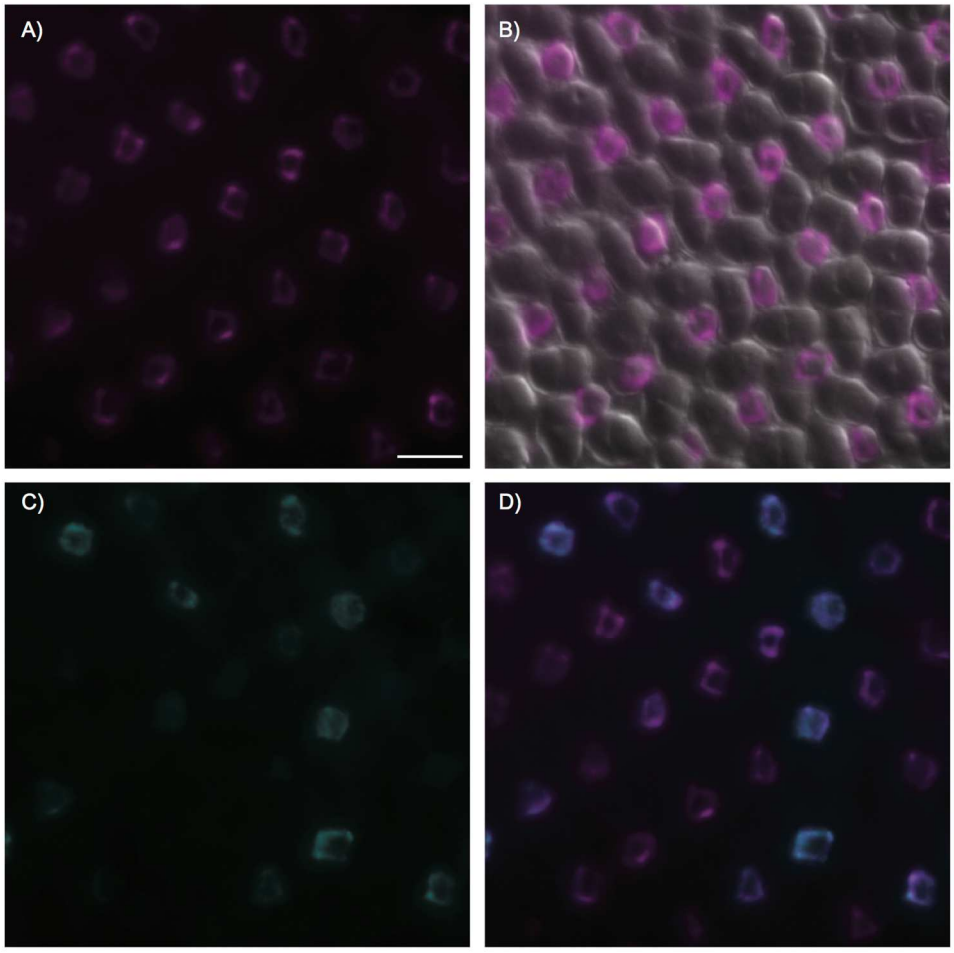


Fig. 2.3: *In-situ* hybridization showing coexpression of *SWS1* and *SWS2B* opsins in the retinas of F2 hybrids. A) *SWS2B* expression in all single cones; Scale bar: 20μm B) Merge of *SWS2B* expression and DIC showing cone mosaic in the retina. C) *SWS1* expression in a few single cones. D) Merge of A and C showing that *SWS1* is coexpressed with *SWS2B*.

One goal of this work is to ascertain the feasibility of a future quantitative trait locus study to map the genomic loci that control opsin expression. The estimated number of genes identified here supports conducting a QTL study to map loci for expression divergence along the *SWS1-SWS2B* axis between *Metriaclima mbenji* and *Aulonocara baenschii*. According to Beavis 1998, one standard deviation difference between the expression levels of the parents gives enough power to detect one QTL in 200 individuals [107]. Using that rule, we speculate that an experiment with 200 F2 hybrids should provide sufficient power

to detect at least multiple, if not all, of the eight genetic determinants for the opsins (11 standard deviations difference between AB and MM for both *SWS1* and *SWS2B* opsins).

2.5.4 Genetic regulation of opsins

Experimental work in humans, rodents and fish have identified that both *cis* and *trans* regulation might be important for opsin expression [87, 117, 118]. Locus control regions (LCRs), which are *cis*-regulatory modules, have been implicated in controlling the differential expression of opsins in primates and zebrafish [93, 119]. Similarly, several transcription factors such as thyroid hormone receptors and retinoic acid receptors have been shown to modulate the expression of opsins [87]. These transcription factors are usually located in genomic regions far from the opsins and facilitate gene expression by binding to specific sites within opsin promoters.

In the present study, we found a statistically significant association of *SWS1* opsin expression with a marker located in the third intron of the SWS1 opsin gene. This marker explains 34% of the variation in expression of *SWS1* between the parents, and points to the possibility of the existence of a *cis* factor contributing to *SWS1* expression differences between the two parents of the cross. This is surprising because previous hybrid crosses in the lab have suggested that differential cone opsin expression in cichlids is not controlled by *cis*-regulatory regions adjacent to the opsin but are controlled by *trans* loci located elsewhere in the genome [63, 64]. Currently, we are fine-mapping this region with microsatellite markers interspersed at regular intervals. Besides confirming the positive association, fine-mapping will help narrow the region enough to determine whether there

is truly a *cis*-factor in the *SWS1* regulatory region or another gene acting in *trans* but closely located near *SWS1*. It should be pointed out that the significantly associated marker only partially explains the variation in expression, so there are likely additional factors acting in *trans* that will hopefully be identified in a full QTL study.

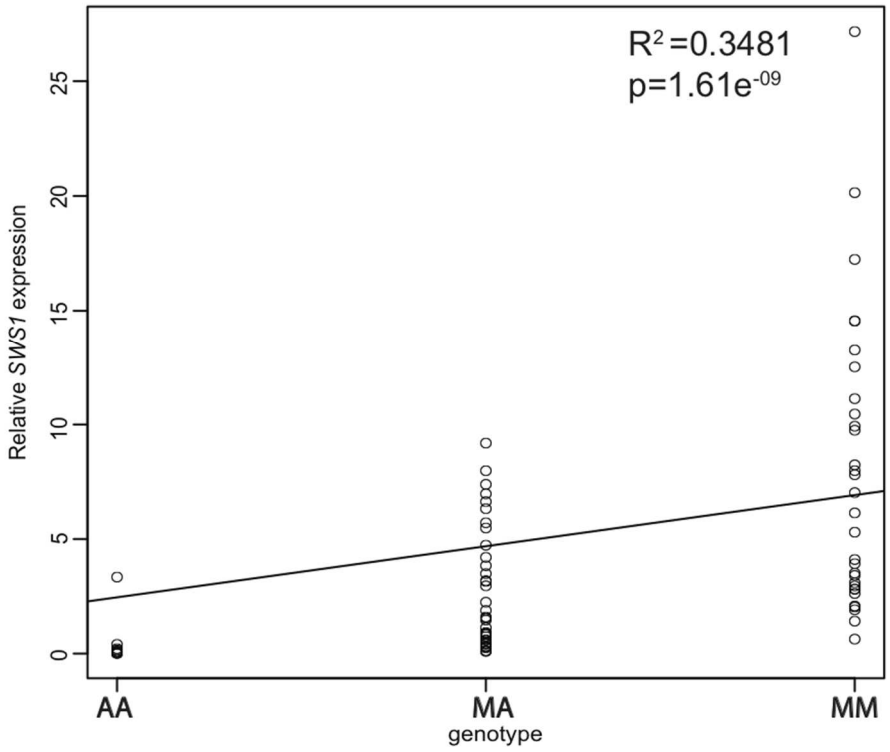


Fig. 2.4: *Cis*-variation for *SWS1* expression. 34% of variation in *SWS1* expression is explained by SWS1_37372 marker located in the third intron of the *SWS1* gene. AA: *Aulonocara baenschi* MA: heterozygote MM: *Metriaclima mbenji.*

2.5.5 Genetic architecture underlying opsin expression

Previous studies have indicated that opsin expression in cichlids varies considerably. Besides intergeneric differences, sister species from Lake Malawi often have dissimilar expression profiles [49, 120]. Furthermore, some species show developmental variation in

opsin expression with different opsins being expressed at different life stages [51]. Adding to the variation in opsin expression are the plasticity-induced changes as shown in the present study. The ease with which opsin expression changes within and between species in cichlids suggests that opsin expression has a relatively simple genetic architecture. Additionally, population sizes in cichlids are small, implying that selection must be strong to overcome drift [121, 122]. Thus, it makes sense that a few genes control a strongly labile phenotype such that this trait can respond to selection rapidly. Results from the current study lend support to this hypothesis with eight loci regulating differential expression of the single cone opsins (*SWS1/SWS2B*) in an intergeneric cross. The lability in opsin expression permits drastic differences in visual sensitivity among closely related Lake Malawi cichlid species. This might have enabled them to rapidly adapt to changing photic environments or to different modes of foraging.

Out of the eight loci estimated for the single cone opsins from the Castle-Wright equation in this study, it is noteworthy that we have already recognized one locus in *cis* to the *SWS1* gene that explains *SWS1-SWS2B* expression divergence. This is interesting because *SWS1* expression has been previously shown to be correlated with foraging [58, 123]. Compared to benthic feeders and piscivores, zooplanktivorous species in Lake Malawi have increased *SWS1* expression. Additionally, behavioral evidence indicates that removal of UV illumination lengthened the duration of finding prey in UV sensitive cichlids [59].

Particularly intriguing is the transgressive segregation of the *LWS* opsin in this cross. Transgressive segregation is a common phenomenon in cichlids and has previously been

observed in hybrid crosses for a variety of phenotypes such as jaw morphology and color patterns [83, 124, 125] [126, 127]. Although caused by several factors, the simplest explanation for transgressive segregation is the presence of two loci with antagonistic effects controlling a trait [128]. It is possible that alleles at opposing loci cancel each other out such that the parental lines show no expression of the trait. Consequently, a combination of parental alleles with effects in the same direction can lead to transgressive expression in the hybrids. If this is indeed the case, future QTL studies will help identify the 'plus' and 'minus' loci for *LWS* expression.

2.6 CONCLUSION

The diversity that characterizes the African lake cichlids makes them an excellent system to address questions regarding adaptive radiation and the evolution of spectral tuning mechanisms. Important future directions include a comprehensive QTL study to identify all the loci responsible for the differential expression of opsins. Knowledge about the proximate genetic mechanisms will provide us a chance to draw inferences regarding the evolution of opsin expression in cichlids, a trait that differs between species and has fitness consequences based on foraging, environmental preference and mate choice. The effects of selection and drift on opsin expression variation could then be readily tested.

Funding

This work was supported by the National Eye Institute at the National Institutes of Health (NIH R01EY024639 to K.L.C).

Author's Contributions

S.P.N made the experimental cross, designed the study, performed the quantitative analysis, microsatellite mapping and plasticity experiments, and drafted the manuscript. B.E.D performed the in-situ experiments. K.L.C conceptualized the study and critically reviewed the article. All authors have read and approved the manuscript.

Acknowledgements

We thank Marissa Schoepp for assistance with collecting broods and dissection of retinas. We thank Daniel Escobar-Camacho for helping with figures, Miranda Yourick and Benjamin Sandkam for valuable suggestions.

<u>Chapter 3</u>: Multiple *trans* factors and one *cis*-regulatory deletion contribute to differential expression of opsins in African cichlids.

In Review under: Nandamuri S, Conte M, Carleton K: Multiple *trans* factors and one *cis*-regulatory deletion contribute to differential expression of opsins in African cichlids. *BMC Genomics*: 2018.

3.1 ABSTRACT

Dissecting the genetic basis of phenotypic diversity is one of the fundamental goals in evolutionary biology. Despite growing evidence for gene expression divergence being responsible for the evolution of complex traits, knowledge about the proximate genetic causes underlying these traits is still limited. African cichlids have some of the most diverse visual systems among vertebrates, with species expressing different combinations of seven cone opsin genes. Using opsin expression variation in African cichlids as a model for gene expression evolution, this study aims to investigate the genetic architecture of opsin expression divergence in this group. Results from a genome-wide linkage mapping of the F2 progeny of an intergeneric cross between two species with differential opsin expression show that opsins in Lake Malawi cichlids are controlled by multiple quantitative trait loci (QTLs). Most of these QTLs are located in *trans* to the opsins except for one *cis*-QTL for *SWS1* on LG17. A closer look at this major QTL revealed the presence of a 691bp deletion in the promoter of the *SWS1* opsin (located 751bp upstream of the start site) that is associated with a decrease in its expression. Phylogenetic footprinting indicates that the region spanning the deletion harbors a microRNA *miR-729* and a conserved non-coding element (CNE) that is conserved in zebrafish and other teleosts. This suggests that the deletion might contain ancestrally preserved regulators that have been tuned for *SWS1* gene expression in Lake Malawi. While this deletion is not common, it does occur in several other species within the lake, suggesting that ancestral polymorphic deletions may contribute to cichlid visual diversity. In addition to expanding our understanding of the molecular landscape of opsin expression in African cichlids, this study sheds light on the molecular mechanisms underlying phenotypic variation in natural populations.

3.2 BACKGROUND

Deciphering the genetic mechanisms underlying phenotypic diversity has long been at the forefront of evolutionary biology research. Much of what we know about the molecular basis for phenotypic variation comes from studies in classic model systems such as yeast and *Drosophila* [68, 129-132]. This work has, without any doubt, vastly advanced our understanding of trait evolution. We now know that the emergence of novel phenotypes can result from both mutations in coding sequences that alter the structure or function of the proteins, or by mutations in the regulatory sequences that modify the spatial and temporal expression patterns of genes. There have been animated debates in the literature on the importance of coding sequence versus regulatory mutations and the relative contribution of *cis* versus *trans* regulatory mutations in bringing about phenotypic change [72, 73, 133, 134]. Although theory predicts that *cis*-regulatory mutations must predominate owing to their highly specific effects and reduced pleiotropy, results from actual studies are often conflicting [135-137]. Adding to the confusion is the fact that some of these studies were performed using inbred wild type and mutant lab raised individuals. Consequently, knowledge about the genetic processes responsible for phenotypic divergence in natural populations has been severely limited. However, recent advances in cost effective sequencing methods allow us to unravel the mechanistic basis for natural adaptation by combining sequence information with traditional quantitative trait loci mapping and candidate gene approaches. Such methods have recently been employed to uncover skeletal appendage patterning in sticklebacks, eye loss in *Astyanax*, feathered feet in pigeons and coat coloration in beach mice [138-141]. Yet a coherent understanding of

the relative roles of *cis*- versus *trans*- regulatory changes requires more studies representing a broader set of traits in naturally-occurring populations.

Vertebrate opsins mainly comprise four main cone opsin families - *SWS1*, *SWS2*, *RH2* and *LWS* [24]. Experimental work in a variety of model systems has identified key factors governing photoreceptor differentiation and opsin expression [86, 87]. These studies have identified that opsins are controlled both by *cis*-factors and *trans*-acting elements. For example, studies in humans have shown that opsin genes are controlled by cis-enhancers located upstream of the genes [142]. Similarly, work in mice has shown that *Thyroid Hormone Receptor Beta* (*THRβ)* transcription factor specifies photoreceptors to a M-cone (*LWS* family) fate [143]. Nonetheless, these studies fail to present a complete understanding of the molecular mechanisms underlying opsin differentiation, simply because model organisms such as mice and humans do not express representatives from all cone families. Owing to gene duplication events, fish typically have a more complex set of cones that are members of the four prototypical vertebrate cone classes [92]. This provides ample opportunity to study genetic factors underlying the specification of opsins that are usually absent in many other vertebrates, such as *SWS1*. In fact, separate studies in trout and zebrafish have identified that *SWS1* opsin is downregulated by *THRβ* and upregulated by *TBX2B* respectively [144, 145]. However, neither trout nor zebrafish have variable *SWS1* opsin expression between closely related species and mechanisms driving the differential expression of *SWS1* opsin largely remain unknown.

African cichlids are a particularly promising system to study the genetic basis of phenotypic diversity [78, 146, 147]. This highly speciose group of freshwater fish inhabit rift lakes throughout Africa, most notably in the three great lakes- Malawi, Victoria and Tanganyika [79]. Incredible bursts of parallel adaptive radiations have occurred in each of these lakes, making cichlids the largest extant vertebrate diversification to date [81, 148, 149]. Just in Lake Malawi, more than 500 species have arisen in the last few million years following multiple repeated colonization and hybridization events [82, 150]. Cichlids are amenable for studying the genetics of adaptive traits for three main reasons. First, different species and even genera of cichlids can interbreed to produce fertile hybrids [50, 83, 151]. This makes mapping of traits for intraspecific and even interspecific comparisons possible. Second, owing to their recent divergence, key polymorphisms are likely to be shared between species [80]. Besides enabling us to find causative mutations, inferences about the evolutionary history of the trait can be easily drawn. Third, well-developed genomic resources are available for cichlids, with complete genomes for the riverine ancestor Tilapia and representative species from each of the three Great Lakes [152].

Within Lake Malawi, cichlids exhibit incredible diversity for a variety of anatomical, physiological and behavioral traits [84, 125, 153, 154]. Closely related species often occupy different habitats, have varied coloration patterns, display numerous feeding strategies and breeding behaviors. For this study, we focus on the diversity of their visual systems, particularly in terms of the genes responsible for color discrimination, the cone opsins. All cichlid species have seven cone opsin genes in their genomes, each maximally sensitive to a different part of the light spectrum [43]. These include three single cone

opsins- *SWS1*(ultraviolet-sensitive), *SWS2B* (violet-sensitive), *SWS2A* (blue-sensitive), and four double cone opsins- *RH2B* (blue-green-sensitive), *RH2A-alpha* and *RH2A-beta* (green-sensitive), and *LWS* (red-sensitive). While the *SWS1* opsin in cichlids is located by itself on LG17, the remaining opsins form two tandem arrays (*SWS2A-SWS2B-LWS* and *RH2B-RH2Aα-RH2Aβ*) separated by 30Mb on LG5 [105].

In Lake Malawi cichlids, large variation in visual sensitivities both within and between species has been shown to be mediated not by coding sequence changes but by alterations in the expression of the cone opsin genes [45, 46, 85]. Adult species from Lake Malawi express only a subset of the seven genes, leading to dramatic differences in their visual sensitivities [49, 50]. Besides being important for foraging, predator avoidance and habitat-preference, vision is key for cichlid fitness because it plays a major role in mate choice and courtship [56, 59-61, 155]. In most species, the males are colorful while females are drab and brown. During breeding season, colorful dominant males display aggressive mating behaviors and get visually selected by choosy females to spawn the next generation [62, 156]. Sister taxa in cichlids often differ in visual sensitivities but not in morphology or behavior. Therefore, it has been hypothesized that the differences in visual capabilities due to variation in opsin gene expression may contribute significantly to the rapid speciation within this group [39, 157].

A genome-wide mapping study using hybrids from an intergeneric cross between *Tramitochromis intermedius* and *Aulonocara baenschi* (TA cross) was previously carried out in our lab to identify genomic regions regulating the differential expression of opsins

[63]. Results from that study identified 11 QTL for 5 cone opsins segregating in the cross- 2 loci each for single cone opsins (S*WS2A* and *SWS2B*- LG5, LG23) and 7 loci for double cone opsins (*RH2B*, *RH2A*, *LWS*- LG1, LG5, LG10, LG15 and LG23) Here, *RH2A*α and *RH2A*β are more than 99% identical and thus have been merged together as *RH2A*. Most of these QTLs are positioned on chromosomes that do not contain the opsin genes they control, indicating that the cone opsins on LG5 are not coordinated by *cis*-regulatory regions adjacent to the opsin, but are regulated by *trans*-acting loci located elsewhere in the genome. Even the QTLs on LG5 are located distantly from the opsin genes, separated by ~ 720 kb, and conservatively classified to be functioning in *trans*. More importantly, since neither of the parental species from the cross had *SWS1* expression, the QTL study was unable to find loci controlling its differential expression.

To find genetic loci controlling the differential expression of *SWS1* opsin, the current study relies on a new hybrid cross between two Lake Malawi species- *Metriaclima mbenji* and *Aulonocara baenschi* (MA cross) [66]. In cichlids, the *SWS1* gene is a single copy opsin on LG17 separated from the rest of the 6-opsin cluster [105]. Previously, we found an association between *SWS1* expression and a micro-satellite marker located in the third intron of the opsin gene, suggesting that *SWS1* opsin could be governed by *cis*-regulation [66]. However, this association was tested with just one marker and only in a subset of the F_2.

Here, we use restriction-site associated DNA sequencing (RADseq) and QTL mapping with markers across the whole genome, using F_2 progeny from the same cross to get a more

complete picture of the genetic mechanisms governing *SWS1* expression. We find three QTLs controlling the differential expression of *SWS1*- two in *trans* on LG14 and LG20, and one located in *cis* to the *SWS1* opsin gene on LG17, indicating that the modulation of *SWS1* opsin expression evolved through changes in both *cis*-regulatory sequences as well as *trans*-acting factors. Also, we identify additional loci segregating for the differential expression of the other opsins (*SWS2B*, *RH2B*, *RH2A* and *LWS*). In addition to presenting a more detailed picture of the genetic architecture of opsin expression in cichlids, this study will lead to a more coherent understanding of the molecular mechanisms contributing to adaptive phenotypic diversity.

3.3 METHODS

3.3.1 Genetic cross

F_2 hybrids generated from an intergeneric cross between two Lake Malawi species *Aulonocara baenschi* and *Metriaclima mbenji* (MA cross) were sampled for this study here [66]. Briefly, hybrid progeny were generated from a mating between a single *A. baenschi* male and a single *M. mbenji* female. A single male and ten females from the F_1 generation were randomly chosen and then mated to generate 350 F_2 progeny (~40 F_2 families). All individuals used in this study were sexually mature (>6 months of age) and raised under laboratory conditions. Sampling was done randomly with respect to sex. Following euthanasia with a lethal dose of buffered MS-222, retinas were dissected and stored in RNAlater at -20C.

3.3.2 Phenotyping opsin expression

Opsin expression in the F_2 hybrids was quantified as described previously [66]. RNA was extracted from dissected retinas and transcribed into cDNA with Invitrogen Superscript III (Invitrogen, NY, USA) following manufacturer instructions. Relative opsin expression from each cDNA sample was measured by real-time, quantitative PCR (RT-PCR) using opsin specific Taqman primers and probes on a Roche Lightcycler 480 according to previously published protocols [43, 53, 158] (*RH2A*α and *RH2A*β genes that are more than 99% identical were combined and noted as *RH2A*). Gene expression was normalized as a percent of total opsin expression within single cones and double cones respectively. Therefore, the fraction, *f*, of *SWS1* expression in the single cones is calculated as:

$$f_{SWS1,\,SC} = T_{SWS1} / (T_{SWS1} + T_{SWS2B} + T_{SWS2A})$$

where T is the transcript level for each of the single cone opsin genes. Similarly, the fraction of *LWS* opsin expression in double cones is:

$$f_{LWS,\,DC} = T_{LWS} / (T_{RH2A} + T_{RH2B} + T_{LWS})$$

3.3.3 RAD-Seq Libraries

To identify SNPs for QTL mapping, restriction-site associated DNA sequencing (RAD-Seq) was performed on F_2 progeny. Reduced representation libraries were constructed by Floragenex Inc. (Portland, OR, USA) according to the protocol of Baird et al., 2008 [159]. High-molecular weight genomic DNA was extracted from fin clips of 157 F_2 hybrids and the 2 F_0 parents using DNeasy blood and tissue kits (Qiagen Inc. CA, USA) and quantified on a BioTek FLX800 fluorometer (BioTek, VT, USA) with Quant-iT dsDNA High-sensitivity kit (Invitrogen, NY, USA). 20ng/ul DNA from each sample was digested with HF-SbfI restriction enzyme, ligated to 95 unique barcode linkers and combined into two

libraries (the second library had 66 F_2 and the two F_0 individuals). Each library was then randomly sheared to an average size of 500bp and size selected between 300-500bp. Following ligation with Illumina adapters and PCR enrichment, both libraries were gel purified and quantified, before being sequenced single-end on an Illumina Hi-Seq 2000 at the University of Oregon Genomics Core Facility.

Raw sequence reads were filtered for quality using the FASTX toolkit. *In silico* analysis for identification of SNPs and determination of genotypes was performed with the built-in wrapper program *denovo_map.pl* in Stacks (version 1.35) as described by Catchen et al., 2011 [160]. Briefly, demultiplexed reads of each individual were assembled into unique loci and heterozygous alleles were identified (allowing for a maximum of 1 mismatch). Then, orthologous stacks from the F_0 parents were matched into a common catalog of loci and differentially fixed SNPs were identified between them. Lastly, the F_2 stacks were compared to the parental stack catalog to infer genotypes at each locus. Conservative criteria were imposed for inferring genotypes from stack loci. Default parameters were modified to require the minimum number of reads to call homozygous genotype in the F_2 to be 10 (*min_hom_seqs*), the minor allele frequency below which a stack is deemed a homozygote to be 1/15 or 0.066 (*min_het_seqs*); the minimum frequency of minor allele to be called a heterozygote to be 2/15 or 0.133 (*max_het_seqs*). Genotypes with minor allele depth between 0.066-0.133 were considered to not have adequate support and were deemed 'unknown'. The RAD loci were filtered further to exclude markers that were not differentially fixed between the F_0 parents. Finally, the consensus sequence of each RAD

marker was aligned to the M_zebra_UMD1_genome assembly using BLAST [161]. Only markers with unique hits in the *M. zebra* genome were included for consequent analyses.

3.3.4 Linkage Map Construction

Genetic linkage mapping was performed in R/qtl following the guidelines described in Broman and Sen [162]. Markers were excluded from subsequent analyses if they were not genotyped in at least 90% of the individuals or failed to satisfy Hardy-Weinberg expectations (after correcting for multiple comparisons). All the RAD-Seq markers were grouped into linkage groups by requiring a maximum recombination frequency of 0.35 and a minimum $\log_{10}$ odds ratio (LOD) threshold of 10. Within each linkage group, markers were then ordered by repeatedly rippling the order of a 7-marker sliding window, ultimately choosing the order with the fewest obligate crossovers. Large-scale changes in marker order were made manually and markers that did not fall into linkage groups were dropped from the analysis. A LOD score for each genotype was calculated in order to identify statistically unlikely events, such as closely placed double crossovers. After excluding such incorrect genotypes, final intermarker distances were estimated using the Kosambi map function [163]. Lastly, linkage groups (LGs) in the constructed map were aligned to the anchored *O. niloticus* (O_niloticus_UMD1) genome assembly to assign to the correct cichlid LG nomenclature [164].

3.3.5 QTL analyses

Quantitative trait loci (QTL) regulating opsin expression were detected following methods in O'Quin, et al., 2012 [63]. QTL analyses were performed in R/qtl using the fully

automated model selection algorithm *Stepwise* that searches for multiple interacting QTL for each opsin [162]. With the stepwise analysis, LOD of association between opsin expression and the genotypes at every marker were computed using forward selection to a model with 10 QTL, followed by backward elimination to the null model of no QTL. Missing genotype data at the putative QTL were taken into account by using Haley Knott regression. The function *fitqtl* was used to calculate the Logarithm of Odds (LOD) score and percentage of variance in opsin expression (PVE) explained by each QTL. Significance thresholds for the QTL were determined from the 95th percentile of the genome-wide LOD scores calculated from 1000 permutations. Markers with the highest LOD score were used to test the effect of each QTL. Interval estimates of the location of the QTL were obtained from 95% Bayes credible intervals (BI) extended to the nearest genotyped markers. For each QTL, candidate genes were estimated by matching the markers flanking the Bayes intervals (BI) to their corresponding regions in the *O. niloticus* (O_niloticus_UMD1) genome [164].

3.3.6 *SWS1* coding sequence analysis

SWS1 coding sequences from *M. mbenji*, *A. baenschi* and *M. zebra* were obtained from Genbank (Accession numbers: *M. mbenji*: HM049223.1, *A. baenschi* GQ422525.1, *M. zebra*: NM_001310074.1). Following translation using the Expasy Translate tool, amino acid sequences from all three species were aligned along with bovine rhodopsin (UniProtKB P02699) using *ClustalW (version 2)* to identify amino acid substitutions that might be functionally relevant [165]. Attention was paid to sites where the substitution

altered amino acid polarity, specifically in the retinal binding pocket of the opsin protein, similar to previous methods [48].

3.3.7 *SWS1* promoter sequencing

The promoter of the *SWS1* gene was Sanger sequenced with overlapping primer sets in both the F0 of the cross as well as in a panel of 17 species from Lake Malawi with previously characterized opsin expression profiles [64]. The primers were designed based on *M. zebra* genome. Primer sequences provided in Table 3.1. DNA from each sample was PCR amplified, size selected by gel-electrophoresis, and sequenced using Applied Biosystems Big Dye Terminators on a ABI 3730xl. For each species, sequences from all the primers were assembled and edited manually in Geneious (*version 11.0.5*) before making comparisons [166]. The *SWS1* promoter sequence alignment file for the 17-species panel has been submitted to Genbank (Accession numbers: XXXX). Since sequencing showed that the deletion was identical across species, 29 additional Lake Malawi species (with known opsin expression) were genotyped for the presence or absence of the deletion by fragment size on an agarose gel [64]. Further, whole genomes from 14 species native to the Malawi catchment area, were obtained from the Wellcome Trust Sanger Institute [167] (12 of the Wellcome genome species are also included in the 29 species Malawi panel, and 2 species- *Aulonocara stuartgranti* and *Astatotilapia calliptera* are not). Sequence data from each species was aligned to the *M. zebra* reference assembly before looking for differences in the *SWS1* promoter. Finally, previously published genomes of cichlids archetypal of Lake Tanganyika (*Neolamprologous brichardi*), Lake Victoria (*Pundamilia nyererei*) and two ancestral riverine species (*Astatotilapia burtoni* and

Oreochromis niloticus) were also examined to look for the deletion in the *SWS1* promoter [152].

Table 3.1: Sequencing primers for the promoter of the *SWS1* opsin gene.

Primer Name	Sequence 5'-3'
SWS1_promoter_3_F	ctggtccaggatgaaaatgg
SWS1_promoter_3_R	gccaaccagctcactgttgt
SWS1_promoter_2_R	ctcagagtccggagaacgtc
SWS1_promoter_1_F	tacctcgcttctcacccact
SWS1_mnpromoter_F	cctcaatgctccaaaaagga
SWS1_mnpromoter_R	cagacaggggccaggtaata

3.3.8 Phylogenetic footprinting assay

DNA spanning from the last exon of *TNPO3* (the gene upstream of *SWS1*) to the first few exons of the *SWS1* gene were obtained from the *M. zebra* genome assembly (LG17: 12796067-12802659). Orthologous DNA regions spanning the same interval were obtained from the genome assemblies of four other teleost fish- Tilapia (*Oreochromis niloticus*, Tilapia_Broad assembly (LG17: 11751000-11756599), medaka (*Oryzias latipes*, NIG/UT MEDAKA1/oryLat2 assembly, scaffold1021: 45422-53836), stickleback (*Gasterosteus aculeatus*, Broad/gasAcu1 assembly, chrUN: 22268595-22274962) and zebrafish (*Danio rerio*, GRCz10/danRer10 assembly, chr4: 13578911-13588818). Comparative sequence analysis was implemented in mVista (accessible at http://www-gsd.lbl.gov/vista/) to perform pairwise alignments of orthologous DNA sequences from multiple species [168]. Sequences were aligned with the global alignment tool MLAGAN and regions were deemed as conserved elements if they showed 50% identity within a 50 bp window for all species except zebrafish [169]. Since the divergence time between *M.*

zebra and zebrafish is nearly twice as much (230 million years) as that of its next distant relative- stickleback (117 million years), the conservation parameters in zebrafish were relaxed and set to 25% identity within the same 50 bp interval (*M. zebra*- medaka: 98 million years, and *M. zebra*- Tilapia: 25 million years) [170].

Transcription factor binding site (TFBS) analysis on the CNE in the promoter of the *SWS1* opsin was performed using the JASPAR CORE Vertebrata 2016 database [171]. Settings were adjusted to shortlist putative transcription factor binding profiles with a relative score of ≥ 0.80. The identity of the microRNA (miRNA) was established using mirBase [172]. Potential target genes for the miRNA were predicted using the computational algorithm *miRanda* according to the methods described in Enright et al 2003 [173]. 3' UTR FASTA sequences for all cichlid genes were obtained from the Tilapia Broad assembly.

3.4 RESULTS

3.4.1 Opsin expression variation in the F2 hybrids

As described in Nandamuri et al., 2017, opsin expression measured by quantitative PCR varied considerably among the 157 F_2 hybrids in the MA cross [66]. Most opsins showed a dominant pattern of inheritance in the F_2 hybrids, with expression of *SWS1*, *SWS2B* and *RH2B* leaning towards levels observed in *A. baenschii* F_0, while that of *RH2A* was closer to levels in *M. mbenji* F_0. Consistent with expression in both the F_0 parents, the F_2 hybrids showed negligible levels of *SWS2A* expression. Expression of *LWS* opsin in the double cones was transgressive with F_2 hybrids showing higher expression than either of the parents of the cross. Since expression of each opsin in the cross is characterized as a

fraction of total expression in single cones (*SWS1*, *SWS2B*, *SWS2A*) and double cones (*RH2B*, *RH2A*, *LWS*) respectively, and since most F_2 hybrids show negligible *SWS2A* expression, *SWS1*and *SWS2B* opsins become inversely correlated.

3.4.2 RAD-Seq Genotyping

After removal of reads with ambiguous barcodes, Illumina sequencing of the two RAD-Seq libraries yielded a total of ~185 million reads for the F_2 hybrids. Around 97% of these reads (~178 million) passed the quality filters and were retained for subsequent processing. This represented a total of ~ 66,500 SbfI restriction sites in the cichlid genome, corresponding to an average read depth of 17x for each F_2 individual. Similarly, after filtering and barcode processing, the average read coverage in the *M. mbenji* F_0 and *A. baenschii* F_0 was 16x (~1 million reads) and 70x (~4.6 million reads) respectively. The low coverage in the *M. mbenji* F_0 was due to the low-quality DNA. In total, Stacks software identified 2370 RAD-Seq SNPs following stringent cutoffs. After filtering to identify differentially fixed SNPs in the F_0, anchoring to a unique location in the *M. zebra* genome and adherence to Hardy-Weinberg equilibrium, we retained a conservative set of 1217 SNPs for linkage map and QTL analysis. The total fraction of missing genotypes across all 1217 SNP markers and 157 F_2 individuals was 4.5% while the frequency of each genotypic class was 26.1% homozygous for *M. mbenji* alleles (MM), 25.4% homozygous for *A. baenschi* alleles (AA) and 48.5% heterozygous (MA).

3.4.3 Linkage Mapping and anchoring to *M. zebra* genome

A high-density linkage map comprising 22 linkage groups spanning 1558.1cM was assembled (Fig. 3.1). The average intermarker distance was 1.3cM and the maximum distance between consecutive markers was 20.9cM. Previous karyotyping work has shown that a haploid cichlid genome is comprised of 22 chromosomes indicating that the linkage map obtained in this study is complete, with one linkage group per chromosome [105]. Also, the length of the map is comparable to previously published linkage maps constructed from other cichlid crosses with closely related species [63, 84, 174]. The linkage map anchored 662,849,923 bp, representing 69.3% of the *M. zebra* genome assembly.

3.4.4 QTL for Single Cone Opsins

Using quantitative trait locus mapping, we detected a total of nine significant QTL for single cone opsin expression- three for *SWS1* (LG14, LG20 and LG17*)*, three for *SWS2B* (LG14, LG20 and LG17) and three for *SWS2A* (all on LG23) (Fig 3.2). However, the three *SWS2A* QTL were likely an artifact of the multiple scan for two reasons: 1) the three loci on LG23 are in close proximity to each other and not separated by at least one genotyped marker. 2) the significant signal is driven by just 10 (of the 157 F_2) individuals showing less than 2% *SWS2A* expression. Therefore, the *SWS2A* QTL were removed from subsequent analysis. Due to the nature of normalization in the single cones and the absence of *SWS2A* expression in most F_2 hybrids, overlapping QTLs were obtained for both the *SWS1* and *SWS2B* opsins. The LOD scores, QTL intervals and percent variance in opsin expression (PVE) explained by each of the QTLs are provided in Table 3.2.

Genome-wide LOD threshold calculated at a significance of P=0.05 is 3.9 for both the *SWS1* and *SWS2B* opsins. For the *SWS1* opsin, the QTL on LG17, spanning a region of 6.2 cM, is the most significant of the three QTL (LOD= 17.97) (Fig 3.3A). Coincidentally, the *SWS1* opsin gene is located within this QTL interval, suggesting that the *SWS1* expression is *cis*-regulated. Additionally, this QTL had the largest effect on *SWS1* expression, accounting for 31.65% of the variance in F_2 hybrid expression. The two QTLs on LG14 (LOD= 6.62) and LG20 (LOD= 8.58) are *trans* QTL explaining 9.8% and 13% variation in *SWS1* expression respectively (Fig 3.3A). The *SWS2B* opsin is located on LG5. Therefore, all three QTL found in this study are located in *trans*. Similar to *SWS1*, LG17 (LOD=17.1) is the largest contributor to *SWS2B* expression, explaining 31.3% of the variance (Fig 3.3B). Around 8.7% and 14.6% F_2 expression variance is explained by the remaining QTL on LG14 (LOD= 5.68) and LG20 (LOD= 9.04) respectively (Fig 3.3B).

Effect plots for *SWS1* show that the two QTLs on LG14 and LG17 are additive with higher *SWS1* expression observed in the F_2 hybrids that are homozygous for the *M. mbenji* alleles (MM>MA>AA) (Fig. 3.4A). However, the QTL on LG20 shows a dominant pattern of inheritance with both heterozygotes and *A. baenschi* homozygotes showing equal levels of *SWS1* expression (MM>MA=AA) (Fig. 13A). Effect plots for *SWS2B* are identical except in the opposite direction (LG14 and LG17: MM<MA<AA; LG20: MM<MA=AA (Fig. 3.4B).

3.4.5 QTL for Double Cone Opsins

The genome-wide LOD thresholds for *RH2B*, *RH2A* and *LWS* opsins are 3.81, 3.91 and 3.78 respectively. For the double cone opsins, we identified a total of six significant QTL: one for *RH2B* on LG16 (LOD= 4.97) (Fig. 3.5A); two for *RH2A* on LG10 (LOD= 6.37) and LG16 (LOD= 8.27) (Fig. 3.5B); and three for *LWS* on LG14 (LOD= 6.46), LG10 (LOD= 15.44) and LG16 (LOD= 4.18) (Fig. 3.5C). All of these QTL are in *trans* to the double cone opsin genes that are located in two separate tandem arrays on LG5. Similar to the single cones, the normalization procedure leads to some degree of autocorrelations between the expression of the double cone opsins. As a result, we obtain overlapping QTLs for all the double cone opsins on LG16, and for *RH2A* and *LWS* on LG10. It is important to note that *SWS1, SWS2B* and *LWS* opsins have one overlapping QTL on LG14 that is independent of the normalization, suggesting that genetic cross-talk exists between the single and double cones (Fig. 3.2).

The LOD scores, QTL intervals and percent variance in opsin expression (PVE) explained by each of the QTLs are provided in Table 3.3. The double cone opsin QTL on LG16 explain 13.56%, 18.2% and 6.67% of the F_2 variance in *RH2B*, *RH2A* and *LWS* expression respectively. While the *LWS* QTL on LG10 is a major contributor accounting for 29.25% of the variance in the F_2 expression, the overlapping *RH2A* QTL only explains 13.61% of its variation. Lastly, 10.65% variance in LWS expression is explained by the QTL on LG14.

Similar to the single cone opsin QTL on LG20, *RH2B* QTL on LG16 shows dominance, with equal levels of expression in F_2 heterozygotes and *A. baenschi* homozygotes

(MM<MA=AA) (Fig. 3.6A). While the expression of *RH2A* opsin on LG10 is additive (MM<MA<AA), heterozygous genotypes at LG16 have lower levels of *RH2A* expression than either of the *M. mbenji* or *A. baenschi* homozygotes, indicating underdominance (MM>MA<AA) (Fig. 3.6B). Lastly, the three loci regulating *LWS* expression on LG14, LG10 and LG16 demonstrate underdominance (MM>MA<AA), additivity (MM>MA>AA) and overdominance (MM<MA>AA) respectively (Fig. 3.6C).

Table 3.2**:** Summary of QTL for single cone opsin expression in the MA cross.

Opsin	LG	location(cM)	LOD	PVE	95% BI (cM)	95% BI (markers)
SWS1	17	39.28	17.97	31.65	38.02-44.21	RAD_33279 to RAD_39835
SWS1	20	81	8.58	13.04	75.64-85.07	RAD_22102 to RAD_52696
SWS1	14	56.81	6.62	9.77	44.96-95.69	RAD_41335 to RAD_2689
SWS2B	17	39.28	17.09	31.3	38.02-44.21	RAD_33279 to RAD_39835
SWS2B	20	81	9.04	14.6	75.64-84.86	RAD_22102 to RAD_8560
SWS2B	14	92.5	5.68	8.72	44.97-95.69	RAD_41335 to RAD_2689

Table 3.3: Summary of QTL for double cone opsin expression in the MA cross.

Opsin	LG	location(cM)	LOD	PVE	95% BI (cM)	95% BI (markers)
RH2B	16	64.19	4.97	13.56	51.27-71.89	RAD_10047 to RAD_19271
RH2A	16	67.5	8.27	18.2	51.27-71.89	RAD_10047 to RAD_19271
RH2A	10	54	6.37	13.61	42.9-54.86	RAD_14019 to RAD_39775
LWS	10	52.91	15.44	29.25	48.02-54.63	RAD_45540 to RAD_1385
LWS	14	52.5	6.46	10.65	44.97-56.81	RAD_41335 to RAD_49138
LWS	16	69	4.18	6.66	51.27-71.89	RAD_10047 to RAD_19271

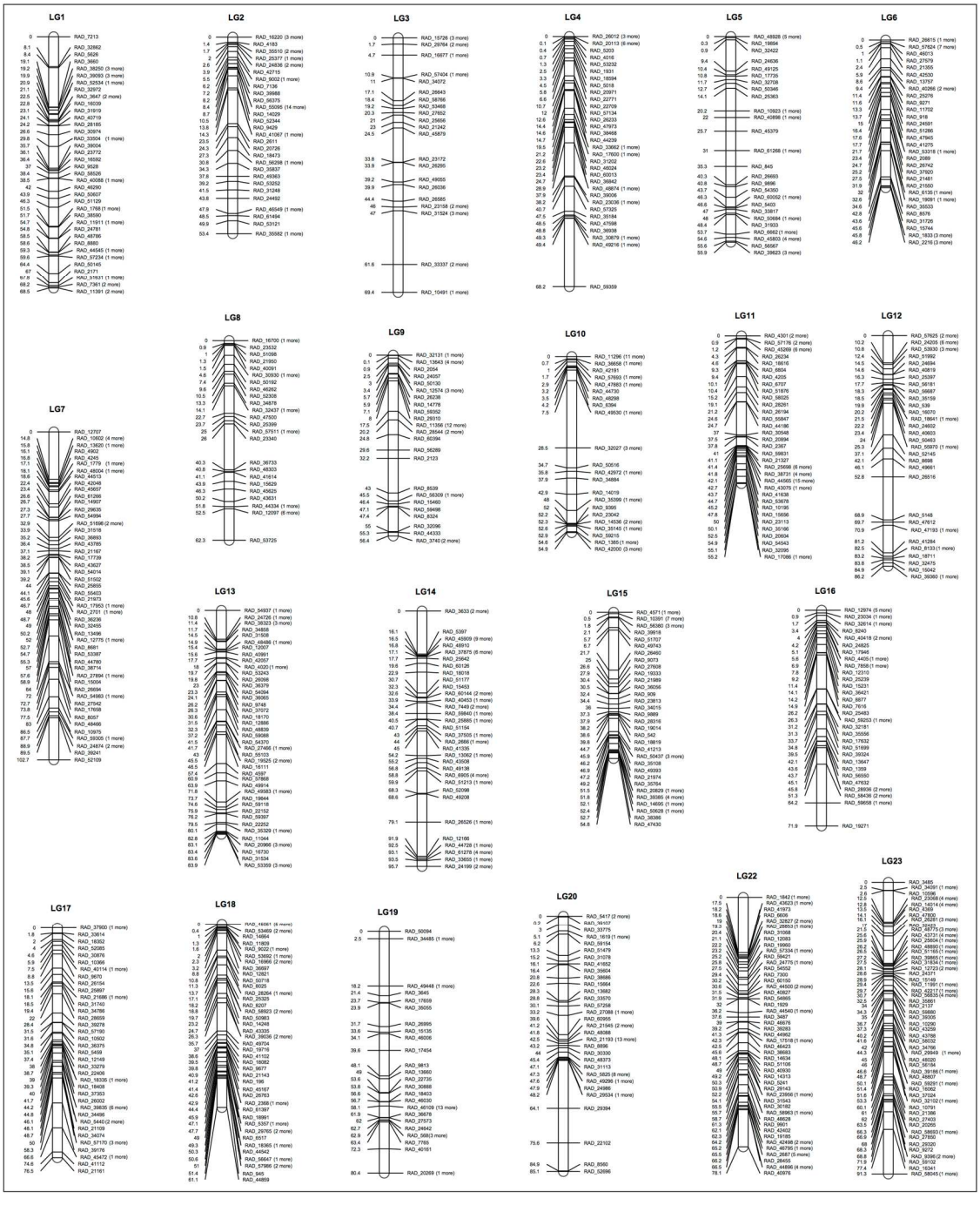


Fig. 3.1: Linkage map constructed from 157 F2 hybrids of an intergeneric cross between two species from Lake Malawi. The map includes 1217 markers in 22 linkage groups spanning 1558.1cM. RAD markers used to construct the map are labeled on the right side of each linkage group. For sites that have more than one marker, the number of additional markers at the location are labeled in brackets. Genetic distances in cM are depicted on the right. The linkage groups are named according to orthologous linkage groups in Tilapia.

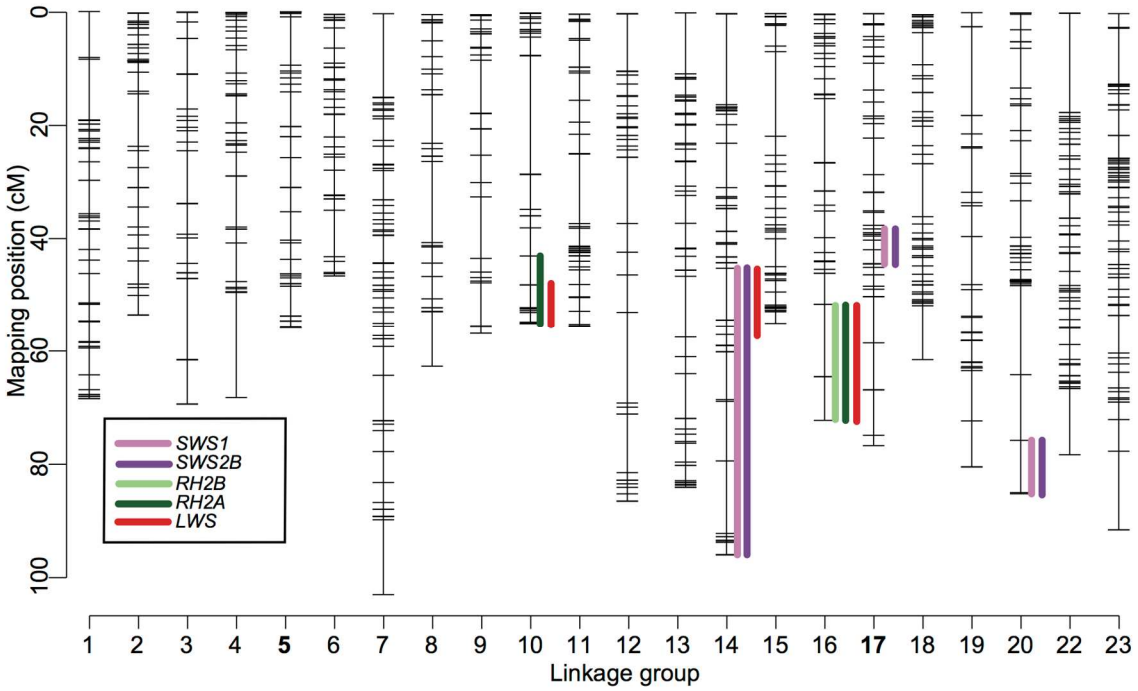


Fig. 3.2: Quantitative trait loci (QTL) mapping in the MA cross. Analysis detected 12 intervals on 5 linkage groups that segregated for the differential expression of opsins between the two F0 species. In the single cones, *SWS1* and *SWS2B* have 3 co-localizing QTLs on LG14, LG17 and LG20. While *RH2B* in the double cones has only 1 QTL on LG16, RH2A and LWS share 2 QTL on LG10 and LG16. Single cones and double cone share 1 QTL on LG14. The opsins in cichlids are present on two chromosomes- LG5 (*SWS2B*, *SWS2A*, *LWS*, *RH2B*, *RH2Aα*, *RH2Aβ*) and LG17 (*SWS1*).

3.4.6 Candidate genes for opsin expression

The QTL intervals on LGs 14, 20, 17, 16 and 10 each map to distinct regions of the *O. niloticus* genome, spanning on average 8 Mb. The *Orochromis niloticus* genome was used since the gene annotations in the Tilapia assembly were complete. Except for the *SWS1* opsin *cis*-QTL on LG17, all the other *trans*-QTL contain numerous genes that potentially could contribute to the differential expression of the opsins (*SWS2B*, *RH2B*, *RH2A*, *LWS*). We have flagged certain genes as the most likely candidates based on their previous

implication with opsin expression in the TA cross and/or association with photoreceptor function in other organisms (Table 3.4).

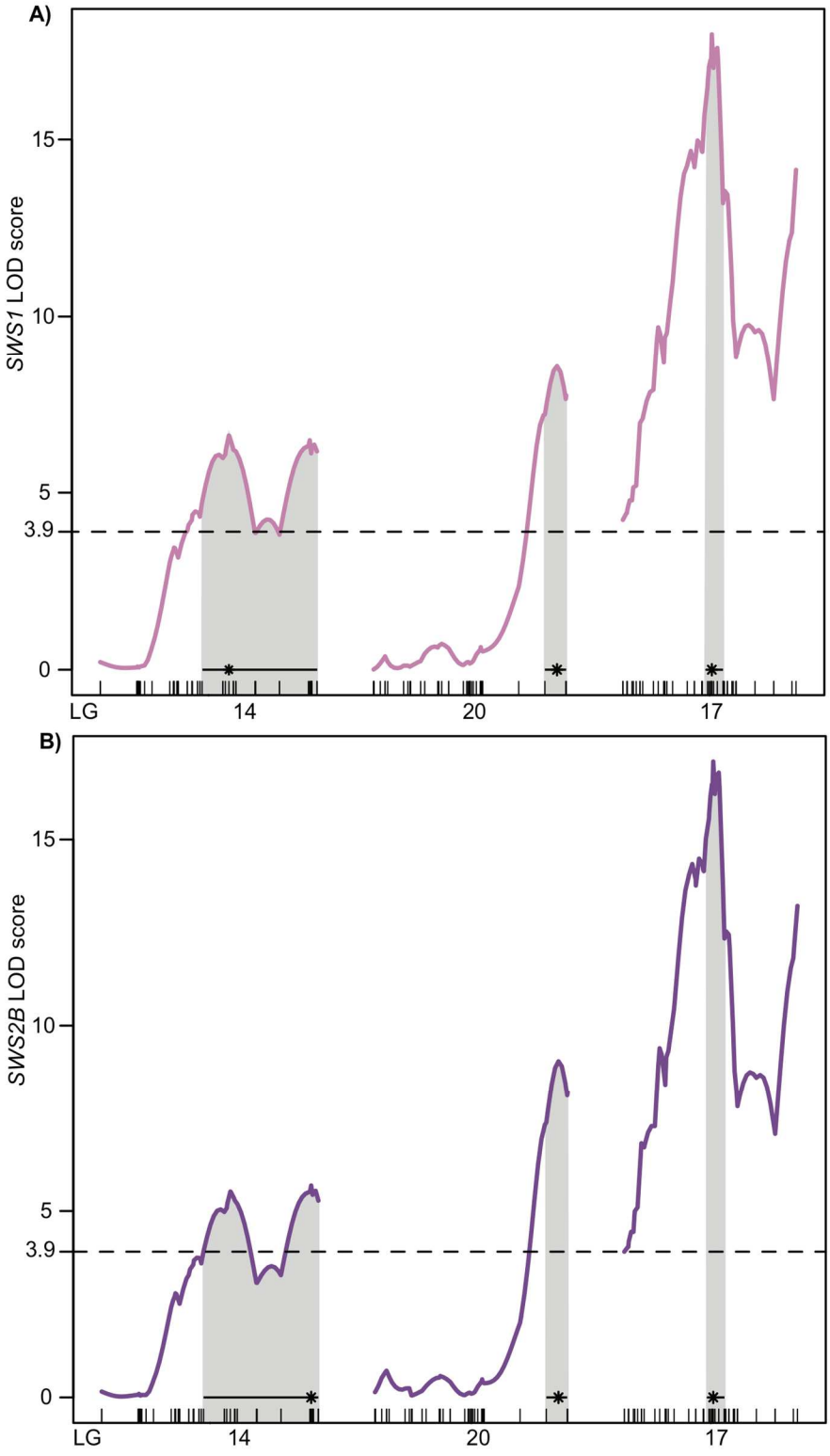


Fig. 3.3: QTL mapping for *SWS1* and *SWS2B* opsins. Horizontal dotted line indicated genome-wide significance threshold for both opsins. The black line at the bottom of each QTL indicates the 95% Bayesian credible interval. The asterisk denotes the location of the most significant marker. A) QTL for *SWS1* expression on LG14, LG20 and LG17.
B) QTL for *SWS2B* expression on LG14, LG20 and LG17.

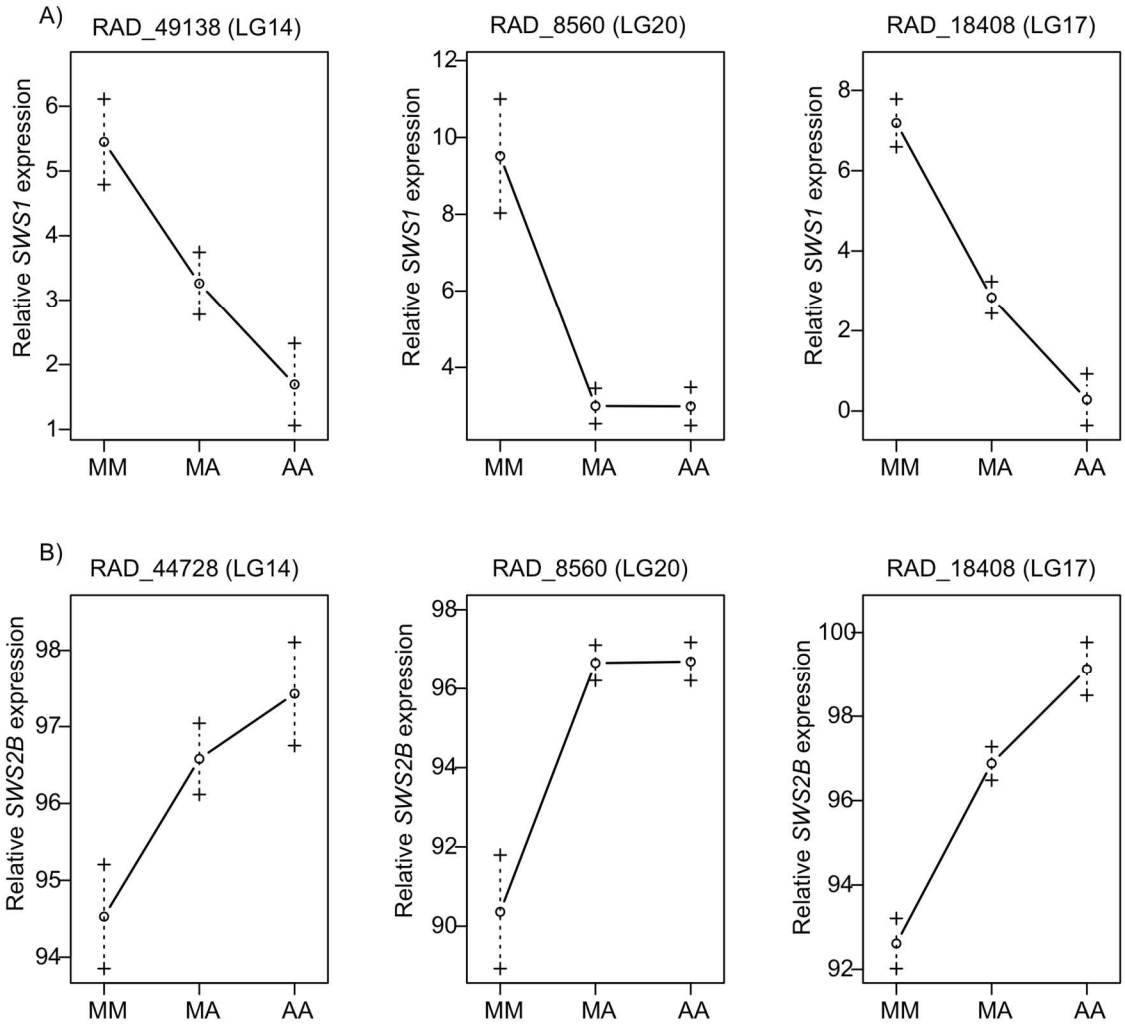


Fig 3.4: Mean opsin expression among genotypic classes at each single cone QTL. Effect plots for each opsin at the marker with the highest LOD score in each QTL. In each plot, MM: *M. mbenji* homozygote, AA: *A. baenschi* homozygote, MA: heterozygote. Error bars indicate standard deviation. A) *SWS1* expression
B) *SWS2B* expression.

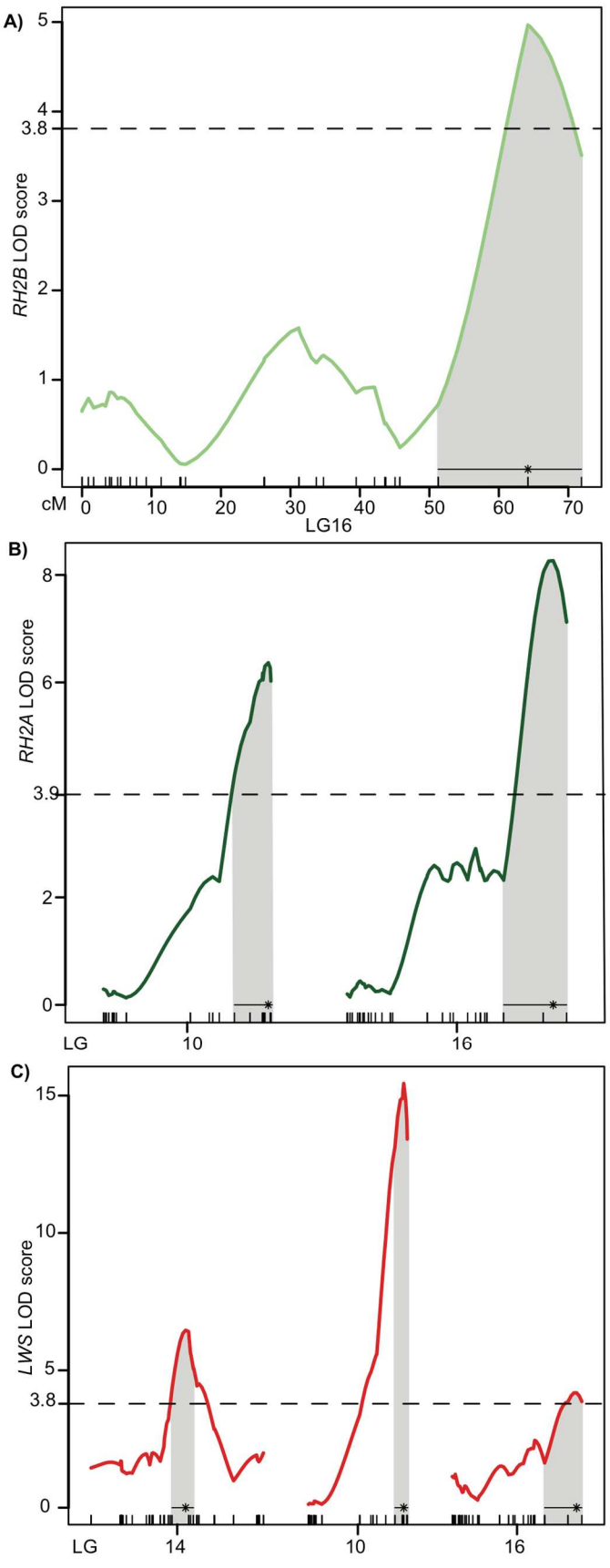

A)
RH2B LOD score
5
4
3.8
3
2
1
0
cM 0 10 20 30 40 50 60 70
LG16
B)
RH2A LOD score
8
6
3.9
2
0
LG 10 16
C)
LWS LOD score
15
10
5
3.8
0
LG 14 10 16

Fig. 3.5: QTL mapping for *RH2B*, *RH2A* and LWS**.** Horizontal dotted line indicated genome-wide significance threshold for all opsins at a significance P=0.05. The black line at the bottom of each QTL indicates the 95% Bayesian credible interval. The asterisk denotes the location of the most significant marker (highest LOD score) in each QTL.
A) QTL for *RH2B* expression on LG16.
B) QTL for *RH2A* expression on LG10 and LG16.
C) QTL for *LWS* expression on LG14, LG10 and LG16.

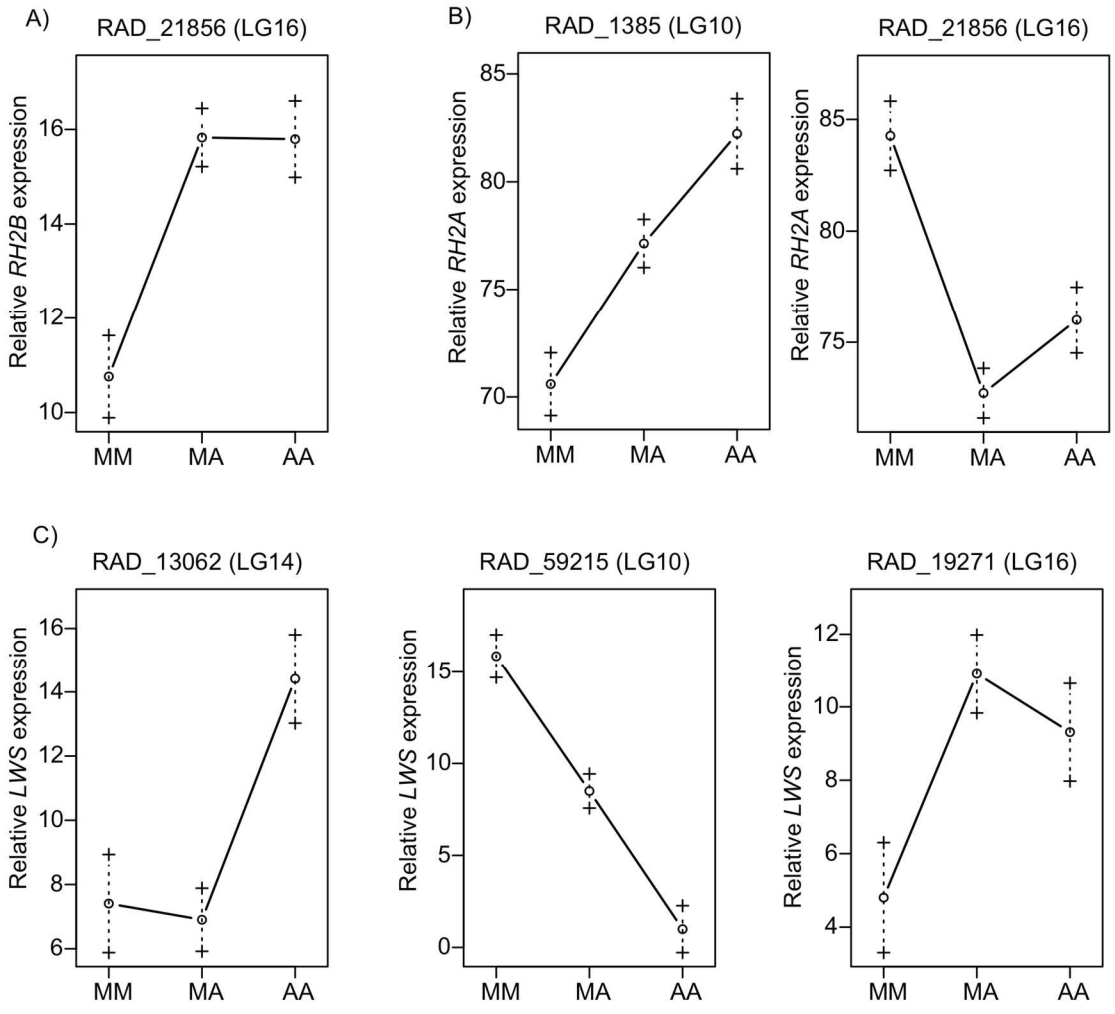


Fig. 3.6: Mean double opsin expression among genotypic classes at each QTL**.** Effect plots for each opsin at the marker with the highest LOD score in each QTL. In each plot, MM: *M. mbenji* homozygote, AA: *A. baenschi* homozygote, MA: heterozygote. Error bars indicate standard deviation.
A) *RH2B* expression
B) *RH2A* expression
C) *LWS* expression.

Table 3.4: Putative Candidate genes for the differential expression of opsins in all QTLs except LG17.

Gene	Location	Accession number	Function	Reference (if available)
Paired box protein Pax-3-B	LG14:3942202	XM_005461712.1	Involved in neural development.	[175]
Zinc finger protein neuro-d4	LG14:4686207	XM_005454306.1	Regulates the proliferation of photoreceptor progenitors	[176]
Max-binding protein MNT	LG14: 14694887	XM_005463727.1	Transcriptional repressor associated with retinitis pigmentosa.	
T-box transcription factor TBX2b	LG10:5759829	XM_005471922.1	UV photoreceptor specification in zebrafish	[144]
Retinal guanylyl cyclase 2	LG10:8655640	XM_003456801.2	Expressed with the cone cells in the retinas of zebrafish	[177]
GPCR Kinase 1	LG10:8655640	XM_003456840.2	Phosphorylates rhodopsin and leads to its deactivation	[178]
Rhodopsin	LG10:8655640	XM_003438995.2	Opsin gene expressed in the rod cells	[179]
Nuclear Receptor Subfamily 1D1	LG20:25027615	XM_005478044.1	Interacts with *NR2E3* and plays critical role in photoreceptor development	[180]

For genes without references, all information was obtained from Genecards (http://www.genecards.org/).

3.4.7 *SWS1* coding sequence comparison between *M. mbenji* and *A. baenschi*

In both grandparents of the MA cross, *M. mbenji* and *A. baenschi*, the 5 exons of the *SWS1* gene encode a 336-amino acid opsin protein that shows extensive sequence similarity to the *SWS1* encoded by *M. zebra* (Fig. 3.7). However, the opsins are not completely identical and differ from each other at 10 amino acid sites. We determined that 2 out of these 10 sites (corresponding to bovine rhodopsin sites 114 and 160) could change the polarity of the amino acid and thus be spectrally tuning, thereby resulting in subtle distinction in the spectral sensitivity of the *SWS1* opsin between the two species.

3.4.8 *Cis*-regulatory elements important for *SWS1* expression

To test for causative regulatory mutations for the differential expression of *SWS1* opsin, we sequenced approximately 2kb of the promoter of the gene in both *M. mbenji* and *A. baenschi* F_0 using overlapping primer sets. We found a 691bp deletion located 751bp upstream of the start codon of the *SWS1* gene in *A. baenschi* (low *SWS1* expression) that was absent in the promoter of *M. mbenji* (high *SWS1* expression) (Fig. 3.8). To test if this region was variable in other cichlid species, we sequenced 17 species from Lake Malawi with previously characterized opsin profiles. Although nine of the 17-species panel had low/negligible levels of *SWS1* expression, none of the species in the panel had the deletion in the *SWS1* promoter. We further analyzed this region for the presence/absence of the deletion in 29 additional species from Lake Malawi by size using gel-electrophoresis. Only one of the 29 species tested, *Placidochromis milomi,* shared the promoter deletion. Additionally, examination of cichlid genomes sequenced by Malinsky et al., 2017 revealed that the same deletion was present in three species- *Aulonocara stuartgranti*, *Placidochromis milomi* and *Trematocranus placodon* [167]. All three of these species have no/low *SWS1* expression. While the *Trematocranus* individual from our lab did not have the deletion, the sample from Malinksy et al., 2017 did, suggesting variation between individuals of a species. This kind of individual variation is not unexpected and has been previously documented in Malawi cichlids [167]. Finally, neither of the representative cichlid species from Lake Tanganyika and Lake Victoria nor the two riverine ancestors had the deletion.

We used mVISTA to perform phylogenetic footprinting of the *SWS1* promoter of *M. zebra* (lacking the deletion) with orthologous regions from several other teleosts (tilapia,

stickleback, medaka), including zebrafish. Our results show that the promoter of the *SWS1* gene bears a high degree of conservation for two well described elements- a miRNA (encoded on the strand opposite to that of the *SWS1* opsin) and a conserved non-coding element (CNE) [181, 182] (Fig. 3.9A, B). Both these elements were located within the region spanning the deletion. *In silico* analysis of the CNE predicted several putative binding sites for transcription factors connected with photoreceptor development in other systems (Table 3.5).

Using mirBase, we ascertained that the predicted sequence of the miRNA is similar to that of zebrafish miRNA, *dre-mi-729*. We used the 3' UTRs of all genes in the Tilapia genome (O_niloticus_UMD1, GCA_001858045.2), to find prospective targets for *miR-729*. Out of the ~28,000 genes predicted in the cichlid genome, we found that the *miR-729* could potentially interact with 1,024 genes. Subsequent filtering of this list based on conserved roles in photoreceptor development and regulation enabled us to identify six top candidate genes that could play key roles in modulating opsin expression through their regulation by *miR-729* (Table 3.5).

Table 3.5: Potential targets of miR-729 and transcription factors (TFs) with putative binding sites in the CNE.

Putative targets of *miR-729*	Potential binding TFs in the CNE
1. *Retinoic acid receptor alpha-A* 2. *Retinoic acid receptor RXR-gamma-B* 3. *Cone-rod homeobox* 4. *Thyroid hormone receptor alpha* 5. *Homeobox protein SIX2* 6. *Thyroid hormone receptor-associated protein 3*	1. *Cone-rod homeobox* 2. *Neural Leucine Zipper* 3. *Orthodenticle Homeobox* 4. *Microphthalmia-Associated Transcription Factor* 5. *Retina and Anterior Neural Fold Homeobox* 6. *Visual System Homeobox* 7. *Ventral Anterior Homeobox*

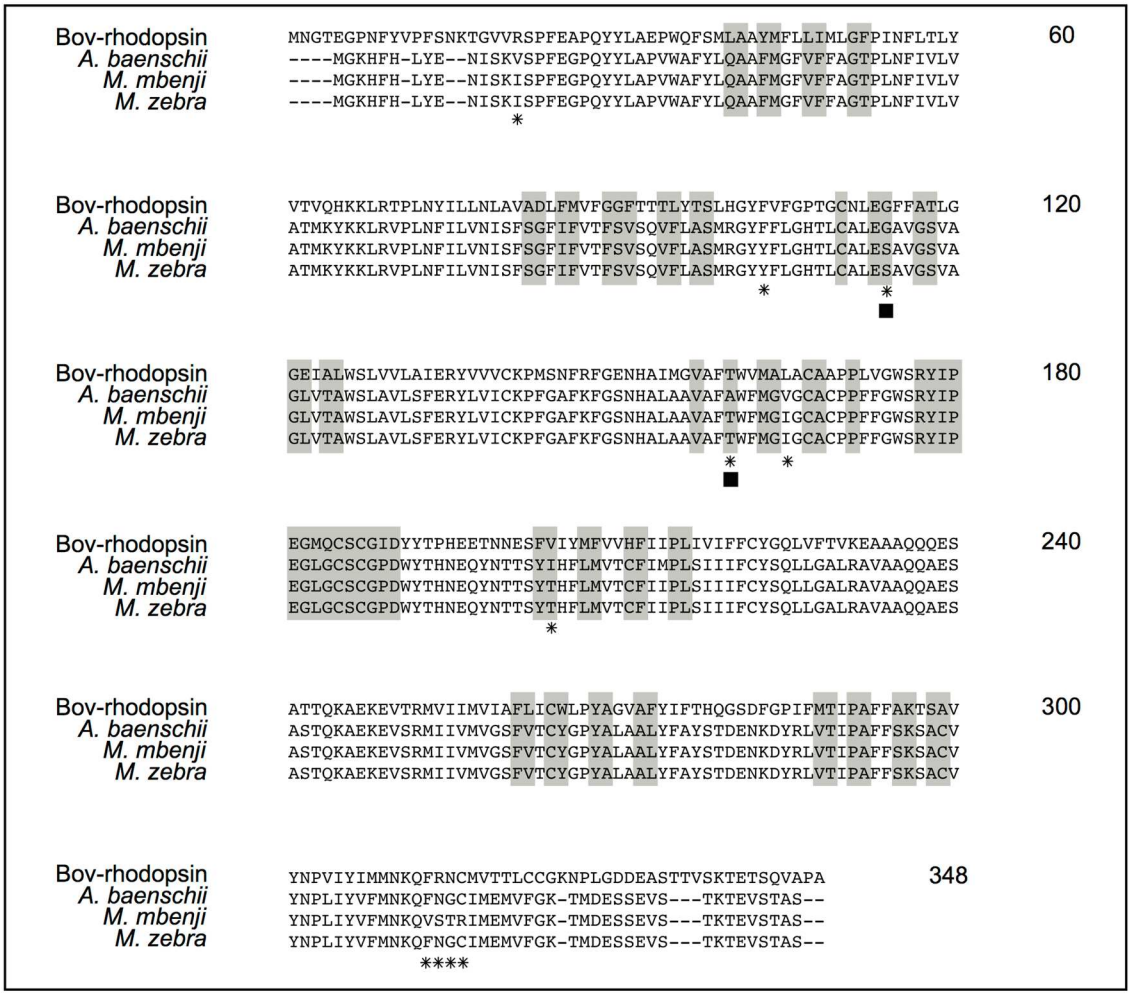


Fig. 3.7: Comparison of coding sequences between *A. baenschi* and *M. mbenji.* Alignment of coding sequence of *SWS1* opsin from the two F0 of the cross, *M. zebra* and Bovine rhodopsin. Gray bars indicate amino acid residues present in the retinal binding pocket that are predicted to be spectrally tuning. 10 amino acids (indicated by asterisks) are different between *M. mbenji* and *A. baenschi*. 2/10 sites (indicated by the black square) represent substitutions that cause a change in the physical properties of the site. These two sites could alter the sensitivity of the *SWS1* opsin in *A. baenschii* relative to *M. mbenji*. Amino acids are numbered according to bovine rhodopsin.

3.5 DISCUSSION

In order to unravel the molecular mechanisms underlying the visual diversity in Lake Malawi cichlids, we utilized an intergeneric cross (*M. mbenji* X *A. baenschi*) and reduced representation sequencing of the F_2 genome to find loci linked with cone opsin expression.

We find a total of 12 QTL regulating the expression of six opsins between the two species, corresponding to five unique genomic loci. It is possible that the modest size of our F_2 population restricted our power to detect additional small QTLs with minor effects and also might have led to an overestimation of the variances explained by the individual QTL [107]. Although the absolute variances for each QTL might be inflated, we hypothesize that the relative magnitudes of the identified QTL for each opsin are most likely accurate. Future fine-mapping with additional F_2 individuals will help provide more exact estimates of the absolute variances explained by each QTL.

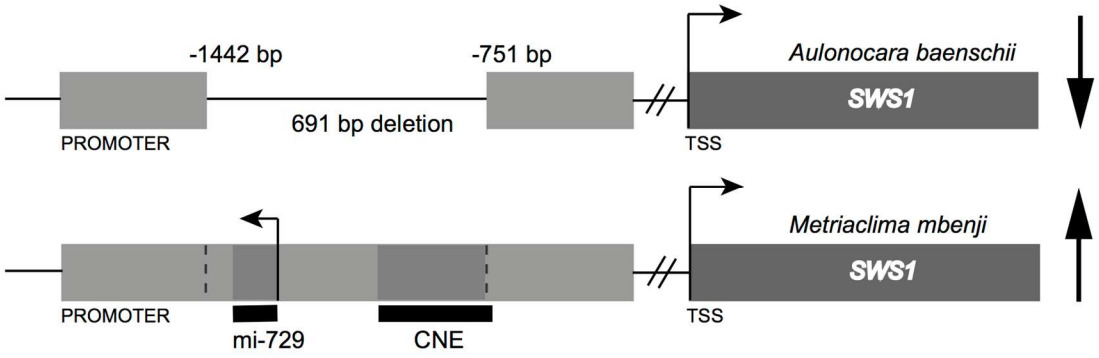


Fig. 3.8: *Cis*-regulatory mutation in the *SWS1* opsin regulates its expression. Schematic representation of the promoter of the *SWS1* gene in both the F0 of the cross. Sequencing revealed a 691bp deletion located 751bp upstream of the translation start site (TSS) of the opsin in *A. baenschii* (low *SWS1* expression) that is absent in the promoter of *M. mbenji* (high *SWS1* expression). The region spanning the deletion comprises a miRNA *miR-729* (transcribed on the strand opposite to that of the opsin) and a conserved non-coding element (CNE). Change in the expression of *SWS1* in both the species is denoted by the black arrow located to the right of the gene.

3.5.1 Regulation of *SWS1* opsin in Lake Malawi Cichlids

QTL mapping revealed three QTLs controlling the differential expression of *SWS1*- two in *trans* on LG14 and LG20, and one located in *cis* to the *SWS1* opsin gene on LG17. The *cis*-

QTL is the largest of the three QTL spanning a 7Mb region on LG17 with the *SWS1* opsin gene located 20bp from the most significant marker. Finding a QTL located in *cis* to the *SWS1* opsin is very surprising because results from this study as well as our previous TA cross suggest that all other opsins are regulated by multiple *trans* acting loci [63].

Comparison of the *SWS1* protein sequences revealed two amino acid differences between the parental species of the MA cross that could potentially alter the maximal sensitivity of the opsin. However, based on evidence from previously site-directed mutagenesis analysis, we theorize that the shift would be very small, shifting the sensitivity of the opsin by just a few nanometers [48]. Thus, the sensitivity of the visual system would be affected negligibly by the coding sequence changes, especially when compared to the substantial shift due to the change in opsin expression from *SWS1* to *SWS2B* (~65nm) [49]. While adult *A. baenschi* have no appreciable expression of *SWS1* in their retinas, *M. mbenji* show high levels of expression [66]. We postulate that the differential expression of *SWS1* opsin is partly due to the *cis*-regulatory mutation in the *SWS1* gene. Specifically, the ~700bp deletion located in the proximal promoter of *A. baenschi* leads to the loss of enhancers needed to drive *SWS1* expression.

Our findings indicate that the region spanning the deletion in the *SWS1* promoter comprise a miRNA *miR-729* and a non-coding element (CNE) that is highly conserved in nearly 250 million years of teleost evolution. Such extensive conservation points to the functional importance of both elements to *SWS1* expression. Exactly how the *miR-729* and CNE modulate the *SWS1* gene and its expression in cichlids is still unclear. miRNA's usually

bind to the 3' UTR of genes and work as post-transcriptional repressors [183]. It is possible that the *miR-729* prevents the misexpression of genes in the wrong cone type. For example, experimental work in mice has shown that *Retinoic Acid Receptor Gamma* (*RXR*γ) suppresses S-cone opsin (homologous to *SWS1* in teleosts) [91]. *RXR*γ is one of the putative targets of *miR-729* in cichlids. We hypothesize that in *M. mbenji* (lacking the deletion), *miR-729* works to keep the levels of *RXR*γ low, so that the *SWS1* cone opsin expression is not repressed. In *A. baenschi*, absence of the *miR-729* due to the deletion can lead to high expression of *RXR*γ, thereby preventing *SWS1* expression. Some of the putative transcription factors that can bind to the CNE in cichlids include upstream genes well-known to be involved with photoreceptor development such as *Cone Rod Homeobox* (*CRX*) and *Orthodenticle Homeobox 2* (*OTX2*). With *A. baenschi* missing these transcription factor binding sites, we postulate that its *SWS1* expression will be low.

It is interesting to note that some of the predicted targets for *miR-729* are the same transcription factors with putative binding sites in the CNE. All cone cells arise from the same pool of retinal progenitor cells and it has been shown that the adoption of a specific cone fate depends on the precisely timed, coordinated and often repeated expression of key transcription factors [184, 185]. It is possible that the spatial and temporal expression of miRNA could act as a negative feedback loop and determine a certain cell-fate decision. Previous work in medaka has shown that the homologous *SWS1* CNE drives the coexpression of both the *miR-729* and the *SWS1* opsin in the same cone cells [182]. This *SWS1* cone-specific expression of *miR-729* in cichlids will be tested in the future with the help of reporter constructs. In order to conclusively prove that the *miR-729* and CNE play

a role in moderating *SWS1* expression, future studies with CRISPR targets against the *miR-729* and specific sites within the CNE will be carried out.

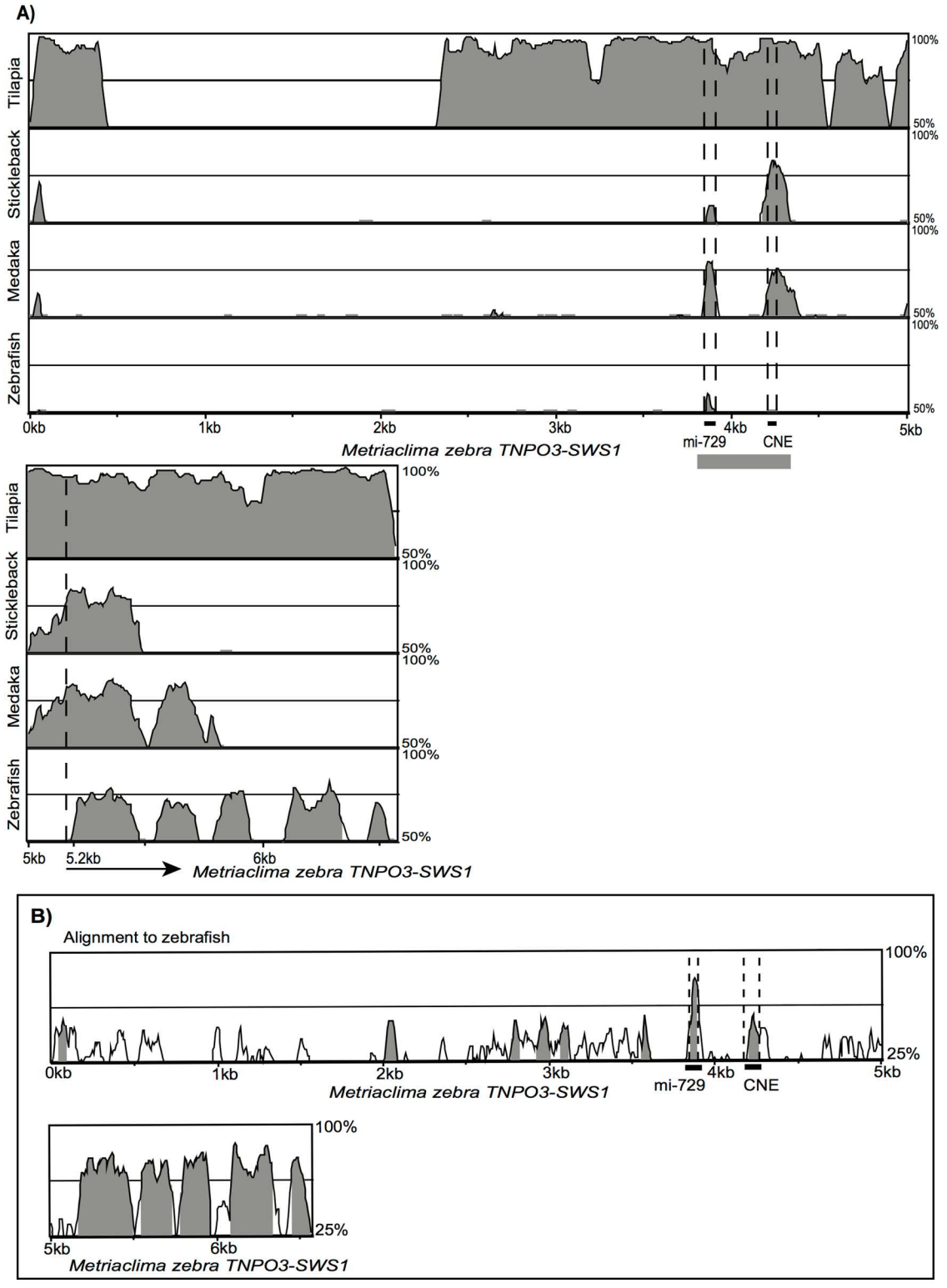


Fig. 3.9: Phylogenetic footprinting of the *SWS1* promoter in teleosts. A) VISTA alignment of the *SWS1* promoter region from tilapia, stickleback, medaka and zebrafish relative to *M. zebra* (X-axis) show that both the mi-729 and CNE are conserved among teleosts. Grey peaks indicate more than 50% conservation in a 50bp sliding windows; the arrow at the bottom indicates the start site of *SWS1* gene. Horizontal grey bar below the X-axis shows the location of the putative deletion in a few species from Lake Malawi. B) VISTA comparison of the same region between *M. zebra* and zebrafish. Grey peaks indicate more than 25% sequence conservation in 50bp sliding windows. Both the mi-729 and CNE are marked by black bars below the horizontal axis.

In trying to determine whether the *SWS1* promoter deletion was common, we found that only 4 of 50 species tested in Lake Malawi had the mutation. All species that had the deletion have low *SWS1* levels but all species with low *SWS1* expression did not have the deletion. This is not surprising because we found two other QTL that influence *SWS1* opsin expression in our cross. It is possible that the differential expression of *SWS1* is achieved by other mechanisms that arose in parallel in various species in Lake Malawi. Interestingly, the species with the deletion were quite disparate and include two species of *Aulonocara,* one species of *Placidochromis,* and one species of *Trematocranus.* This suggests random sorting of the deletion into different lineages as occurs for ancestral polymorphisms. None of the species outside the lake, particularly species from Lake Victoria (*P. nyererei*) and Tanganyika (*N. brichardi*) and the riverine species (*O. niloticus* and *A. burtoni*) had the deletion indicating that the deletion in the *SWS1* promoter arose within Lake Malawi and is likely being shared either by hybridization or sorting through the flock as standing genetic variation [150, 186]. Future work looking at whole genomes from various species in each of these lakes will enable us to identify if other species carry the deletion.

3.5.2 Opsin expression in cichlids is controlled by multiple, overlapping QTL

With the exception of one *cis*-QTL for *SWS1* on LG17, our results indicate that most opsins in cichlids are controlled by *trans* loci located in genomic regions distant from the opsins. Being upstream of the immediate genes regulating the trait, *trans* changes are often thought to be inefficient because of the widespread and often negative pleiotropic effects achieved by simultaneously affecting multiple downstream targets [187]. We speculate that in cichlids, *trans* mutations might actually be favored for two main reasons. First, evidence from other organisms points to the existence of Locus Control Regions (LCRs) that modulate the expression of downstream opsin tandem arrays [93]. These LCRs are usually driven by multiple *trans* acting factors operating as switches. By virtue of its sequestered location (LG17), *SWS1* opsin need not be driven by a LCR and thus its promoter might be susceptible to accumulating *cis*-changes. Second, closely related species in Lake Malawi often exhibit coordinated differences in the expression of multiple opsins [120]. Such synchronized changes can be more easily achieved by a single *trans* change influencing multiple downstream opsins as opposed to individual *cis* mutations in the opsin promoters.

The QTL for single cone (LG14, LG20 and LG17) and double cone opsins (LG10 and LG16) are co-localized to the same chromosomal regions. Although this might be a direct outcome of the normalization procedure used in the study, previous in-situ hybridization work both in the hybrids of the MA cross as well other Malawi species have shown that in conditions of altered light spectra, both the single cone and double cone opsins are co-expressed in their respective cell types in certain parts of the retina [55, 66, 103]. Therefore, it is not unreasonable to hypothesize that at least a few loci controlling their expression might be shared as well. Co-mapping of a QTL for LWS with 2 QTLs for *SWS1* and *SWS2B*

opsins, independent of the normalization, is striking. Such genetic coupling between the single and double cone opsins hints at pleiotropic effects, although at this stage multiple linked genes cannot be ruled out. Future fine-mapping of the shared QTL interval on LG14 will help resolve this issue.

Previously, Castle-Wright estimates based on gene expression levels in the three generations of the cross (F_0, F_1 and F_2) suggested that ~8 genes govern differences between *SWS1/SWS2B* opsin expression in the MA cross [66]. In the current study, we were able to detect only 3 of 8 loci for these opsins. Failure to recognize additional loci might be due to the complex architecture of the single cone opsins including multiple genes with pleiotropic and epistatic effects along with dominance. For the green-sensitive opsins in the double cones (*RH2A* and *RH2B*), we were successful in identifying 3 QTL on LG10 and LG16. This was more than what had been previously predicted from the Castle-Wright estimator. The *LWS* opsin exhibits transgressive expression in the F2 generation of this hybrid cross, which we previously theorized was due to the presence of two loci with opposing effects in each parent [66]. Subsequent segregation of the alleles with effects in the same direction leads to F2 offspring with high expression levels. Consistent with our hypothesis, in the present study we found 3 QTL for *LWS* opsin with divergent effects. While the inheritance of *M. mbenji* alleles in the largest QTL on LG10 increases *LWS* expression, the *M. mbenji* alleles seem to decrease expression in the other QTL on LG14 and LG16. Failure to find the fourth QTL for *LWS* as projected by the Castle-Wright estimator might be due to the limitations discussed above.

Together with our previously published QTL generated from the TA cross, we now have a total of 23 QTL regulating the differential expression of all six cichlid cone opsin classes. Of the 23 QTL, two (*RH2A* and *LWS* on LG10) replicate between the two studies. Also, the putative QTL for *SWS2A* expression on LG23 (which we discarded in this study, for reasons- see Results) matches the QTL of the same opsin from the TA cross. The common parental species between the two crosses, *A. baenschi* might be the source of these replicate QTL. Nevertheless, we postulate that QTL results from both the crosses can be combined for two main reasons- First, different species in Lake Malawi all diverged from a recent ancestor less than a million years old [80]. This recent evolutionary history resulted in the maintenance of a high degree of homology between the genomes of different species. Most cichlid species have similar genomic structure, with contiguous stretches of nearly identical sequence. Second, standing genetic variation, mostly as ancestrally-inherited polymorphisms, seems to be an important source for genetic diversity in cichlids [186]. Repeated selection of the same allele independently by different species can lead to shared genetic mechanisms. Currently, fine-mapping and identification of candidate gene (and causative mutation) in the QTLs from the TA cross is in progress in the lab. It will be interesting to see if the same genes can explain the variation in opsin expression in the MA cross as well.

3.5.3 Candidate genes for differential expression of opsins

All the *trans* QTL identified in this study correspond to discrete regions in the genome and contain numerous candidate genes. Although the list provided in Table 3.5 is drastically condensed, it is the first step towards identifying suitable candidates and also provides a

point of reference for our fine-mapping efforts. There is the possibility that the differential expression of the opsins in these regions is being regulated by other factors such as microRNAs, non-coding RNAs and epigenetic elements that we have not considered here.

One of the candidate genes for *LWS* expression is *T-box Transcription Factor TBX2B*, located in the QTL on LG10. In zebrafish, this gene has been implicated in promoting the production of *SWS1* cone cells by repressing the rod differentiation pathway [144]. Another good candidate is *NEUROD4*, a cell cycle regulator gene on LG14 that promotes photoreceptor fate, while inhibiting other retinal cell fates in zebrafish [176]. Other genes connected to photoreceptor function thereby making them suitable candidates include *GRK1* on LG10 (phosphorylates the rod opsin leading to its inactivation) [178], and *NR1D1* on LG20 (affects the survival of the S-cones; mutations in this gene lead to increased sensitivity to blue light in humans) [180].

3.6 CONCLUSION

In this study, we have shown that both *cis* and *trans* regulatory changes contribute to the divergence in photoreceptor sensitivity in African cichlids. Additionally, we uncover a regulatory deletion in close proximity to the start site of the *SWS1* opsin that likely aids its expression variation in these fishes. This is the second regulatory deletion that we have identified as possibly playing a role in cichlid opsin expression (see [64]) suggesting that indels may be an important source of variation and might contribute to cichlid explosive radiation.

Expression of *SWS1* opsin in Lake Malawi cichlids is tied to their visual ecology, having been previously linked to foraging and predatory behavior [58, 59]. Zooplanktivorous cichlids and algal browsers in Lake Malawi usually have higher *SWS1* expression than species that do not rely on plankton for their nourishment. Furthermore, behavioral evidence in cichlids with *SWS1* expression shows that the duration of finding prey is significantly increased under illumination conditions devoid of ultraviolet input. Finding the *cis*-regulatory deletion and other loci controlling its differential expression serves as the first step towards understanding the adaptive origins of this trait.

Funding

This work was supported by the National Eye Institute at the National Institutes of Health (NIH R01EY024639 to K.L.C).

Author's Contributions

S.P.N generated the hybrid cross, performed the QTL experiment and analyses, and drafted the manuscript. M.A.C performed the bioinformatic analysis pertaining to the study and reviewed the manuscript, K.L.C conceptualized the study and critically reviewed the article. All authors read and approved the final manuscript.

Acknowledgements

We thank Kelly O'Quin for assistance with map construction and QTL analysis. We thank Marissa Schoepp for helping with the retina dissections. We thank Daniel Escobar-Camacho for helping with figures, Miranda Yourick and Benjamin Sandkam for valuable suggestions on the manuscript.

Chapter 4: Adult plasticity in African Cichlids: Rapid changes in opsin expression in response to environmental light differences.

Previously published under: Nandamuri SP, Yourick MR, Carleton KL: Adult plasticity in African cichlids: Rapid changes in opsin expression in response to environmental light differences. *Mol Ecol* 2017, 26(21):6036-6052.

4.1 ABSTRACT

Phenotypic plasticity allows organisms to adapt quickly to local environmental conditions and could facilitate adaptive radiations. Cichlids have recently undergone an adaptive radiation in Lake Malawi where they inhabit diverse light environments and tune their visual sensitivity through differences in cone opsin expression. While cichlid opsin expression is known to be plastic over development, whether adults remain plastic is unknown. Adult plasticity in visual tuning could play a role in cichlid radiations by enabling survival in changing environments and facilitating invasion into novel environments. Here we examine the existence of and temporal changes in adult visual plasticity of two closely related species. In complementary experiments, wild adult *Metriaclima mbenji* from Lake Malawi were moved to the lab under UV-deficient fluorescent lighting; while lab raised *M. benetos* were placed under UV-rich lighting designed to mimic light conditions in the wild. Surprisingly, adult cichlids in both experiments showed significant changes in the expression of the UV-sensitive single cone opsin, *SWS1*, in only three days. Modeling quantum catches in the light environments revealed a possible link between the light available to the *SWS1* visual pigment and *SWS1* expression. We conclude that adult cichlids can undergo rapid and significant changes in opsin expression in response to environmental light shifts that are relevant to their habitat and evolutionary history in Lake Malawi. This could have contributed to the rapid divergence characteristic of these fantastic fishes.

4.2 BACKGROUND

Phenotypic plasticity is the ability of an individual genotype to produce multiple distinctive phenotypes in response to varied environmental conditions [5, 7, 188]. The environment presents certain triggers, whether biotic, such as social interactions between potential mates or predators, or abiotic, such as seasonal changes in temperature or the current light environment. Additionally, these triggers are not fixed, but in a constant state of flux, fluctuating variably throughout the lifetime of the organism. Responding appropriately to these heterogeneous environmental cues is linked to organismal fitness and is crucial for executing fundamental tasks such as foraging, mate-choice and predator avoidance. Thus, phenotypic plasticity can increase the persistence of a species in variable environments and can also facilitate the ability to adapt to novel niches [8, 189].

Historically, phenotypic plasticity has been overlooked as being important to adaptive evolution because it could possibly shelter the genotype from the effects of selection [6, 190]. However, recent theoretical modeling and experimental research has revealed that under the right conditions, environmentally induced plasticity can promote diversification and speciation events, such as in adaptive radiations [16, 191-193]. This could be accomplished in several ways, some of which include allowing the immediate exploitation of different niches in a novel environment, producing a rapid response to selection pressures in one generation, and enabling the repeated evolution of traits suited to different environments [194, 195]. Phenotypic plasticity is ubiquitous in nature and has been observed for several ecologically relevant traits in phenotypically diverse lineages such as social behavior in Hymenopterans [196], courtship behavior in sticklebacks [197], life

history traits in nematodes [198], color morphs in spiders [199] and limb length in Anoles [200]. In the present study, we concentrated on visual system plasticity, mainly as changes in the opsin gene expression in response to environmental lighting that could enhance an animal's visual function and ultimately its fitness.

Photoreceptors lead the visual response, being specially harnessed to capture light. The absorption of photons by the photoreceptors starts a signal transduction cascade resulting in neural output [20]. As such, the ability of an organism to detect and respond to light will depend on the absorption properties of the visual pigment residing within the photoreceptors [27, 201] . Indeed, it has been extensively shown that subsets of vertebrate photoreceptors have spectral sensitivities that closely match the most abundant wavelengths in the surrounding environment [26, 202, 203]. This spectral "tuning" of the photoreceptors to ambient light is ecologically relevant and serves to enhance contrast detection in visually driven tasks [36]. Additionally, variation in the properties of the photoreceptors due to differences in the spectral range of environmental light has been known to foster population differentiation and speciation [35, 39].

Visual sensitivities of animals can be adapted to their environmental light via genetic modifications and/or phenotypic plasticity. Genetic changes such as structural replacements in specific amino acids within the opsin protein moiety shape the sensitivity of the visual pigments [21, 204]. These genetic modifications usually take place slowly, accumulating gradually over evolutionary time [205, 206]. On the other hand, plasticity induced changes fine tune visual acuity much more rapidly, within the lifetime of the

organism [207, 208]. Some of these physiological mechanisms include seasonal reversals in chromophore usage in connection with migration behavior [209] or alterations in the dietary availability of carotenoids [210, 211].

One other molecular mechanism for altering spectral sensitivity that can be both genetically determined and environmentally induced, thereby having the potential to act across varied time scales, is changes in opsin gene expression [50, 51, 212]. Across species, opsin expression variation in African cichlids causes significant differences in spectral sensitivities between closely related taxa [45, 49]. This differential opsin expression results in closely related species with huge differences in their visual capabilities and aids cichlid diversification across all major axes such as habitat, foraging and mating preference [56, 59-61]. Within a single species, opsin expression changes can be hardwired ontogenic events, often correlated with changes in habitats or foraging behaviors [51]. For example, cone opsins in riverine Tilapia convert their expression from shorter to longer wavelengths as the fish grow and reach adulthood. This kind of developmental progression has also been observed in other fish species such as salmon [213], goatfish [214], eel [215], trout [216], flounder [217] and black bream [218].

In addition to the genetic mechanisms described above, several studies have examined whether there can be phenotypic plasticity in opsin expression in response to changes in environmental light conditions. However, results from these studies are often conflicting; with some organisms showing plastic changes while others have little/no plasticity whatsoever. When reared under different light regimens, developmental plasticity in opsin

expression was observed in guppies [219] and black bream [31]. Contrary to these results, rearing conditions had no effect on opsin expression in cave-dwelling mollies [220] and South American cichlid *Aequidens pulcher* [221]. Although not tested during development, adult sticklebacks showed minor plasticity in opsin expression [222]. So far, the only known system that shows both developmental and adult plasticity in opsin expression is the bluefin killifish (*Lucania goodei*) that occurs in a wide-range of lighting conditions, from tannin swamps to clear freshwater [212, 223]. Two separate experiments have documented that killifish show pronounced developmental and adult plasticity in opsin expression [224, 225]. Larval killifish reared in tea-stained water have higher *RH2* and *LWS* expression, while fish from the same brood reared under clear water have increased expression of *SWS1* and *SWS2B* opsins. Similar changes were also observed in swamp and clear spring adult fish when placed in reciprocal water conditions. Importantly, these plastic changes have been observed in as little as three days.

The cichlids of East African rift lakes are one of the fastest speciating vertebrates with over 500 species from Lake Malawi having evolved from a common ancestor less than a million years ago [80, 81, 146]. Due to their recent evolutionary history, all cichlid species have nearly identical sequences for each of their seven corresponding cone opsin genes (*SWS1*, *SWS2B*, *SWS2A*, *RH2B*, *RH2Aβ*, *RH2Aα* and *LWS*) [46]. Large differences in visual sensitivities among cichlid species are achieved by changing the complement of expressed opsin genes [49, 85]. Adult cichlids from Lake Malawi express a subset of the seven cone opsin genes in their retinas, leading to sizeable differences in the spectral properties of their photoreceptors.

In cichlids, modification of spectral sensitivity due to changes in opsin expression are coordinated both by genetic and environmental factors. Multiple hybrid crosses between cichlid species with differing opsin expression patterns have revealed that the largest opsin expression differences are controlled by a few loci [50, 63, 66]. Additionally, environmentally induced plasticity plays an important role in tuning visual sensitivity during development. Previously, Lake Malawi cichlids reared under fluorescent light conditions in the lab expressed small but significantly different opsin expression profiles than their wild counterparts [53]. Further, differences in opsin expression were observed when individual broods were split and raised under different lighting conditions [54]. Also, developmental plasticity was observed in cichlids with increased *LWS* expression in dorsal retinas when fish were raised under upwelling fluorescent light, [55]. Although these earlier studies were successful in establishing that the visual system in cichlids is developmentally plastic, two crucial questions remain unanswered: 1) Can plastic changes in opsin expression occur in adulthood or are they restricted to early development? 2) What are the temporal dynamics of the plastic change in opsin expression? In other words, how quickly can adult individuals modify their opsin expression in response to an altered environmental cue?

The current study has three main goals. First, we want to examine if adult cichlids exhibit plastic changes in opsin expression due to differences in lighting conditions. If visual tuning by plasticity is only restricted to development, it could limit the ability of cichlid species to adapt to changing light environments at later life history stages. Second, we want

to monitor the tempo of changes in opsin expression to determine how swiftly plastic changes are observed. If the plastic response to an altered light regime takes too long to come into effect, it might adversely affect the optimum functioning of the visual system in the new light environment. Third, we want to compare opsin expression relative to a cone-specific housekeeping gene with opsin expression normalized to the total opsin expression. Many of our previous studies have compared opsin expression relative to total pool of opsins within cone type [50, 123]. While this enables comparisons of opsin ratios, it obscures which opsin is actually changing in expression due to differences in illumination. To study plasticity, we set up two complementary experiments: 1) we took adult wild-caught cichlids from a single population and housed them in narrow spectrum lighting devoid of UV wavelengths (Natural to Narrow); and 2) we placed lab-reared adult cichlids in high intensity UV-rich broadband illumination designed to mimic light in the natural lake environment (Narrow to Broad). Results from this study give us the first insights that cichlids with short wavelength sensitivities can indeed respond on a rapid time scale to light changes.

4.3 MATERIALS AND METHODS

4.3.1 Species selection

Two species from a single genus were chosen for the experiments. *Metriaclima mbenji* were used for the Natural to Narrow study and *Metriaclima benetos* were used in the Narrow to Broad experiments. These sister species are thought to be less than 100,000 years divergent to each other and to *Metriaclima zebra*, which is the representative cichlid species from Lake Malawi [97, 152]. Both species occupy similar ecological niches in the

lakes with identical foraging, mating and breeding behaviors. Most importantly, adult wild-caught individuals of both species have the same opsin expression profiles in their retinas - *SWS1* and minor levels of *SWS2B* expression in the single cones; and *RH2B* and *RH2A* expression in the double cones [49]. When raised under different lighting conditions, a plastic response in opsin expression has been previously demonstrated in *M. mbenji* [53]. Since *M. benetos* is a sister species to *M. mbenji,* it was presumed that they would also be plastic. Additionally, developmental plasticity has been observed in *M. benetos* when reared in the lab [51]. For the Natural to Narrow study, adult *M. mbenji* individuals from a single population (n=49) were collected in 2015 from Mbenji Island (13º 26'23. 50"S 34º29'35"E) and housed in the lab for the duration of the experiment. Adult *M. benetos* (n=47) used for the Narrow to Broad study were multiple generation progeny produced from wild-caught stocks collected from Mazinzi reef in 2008 (14 º 5'33"S35 º 1'94"E).

4.3.2 Light environments

The light environments were selected to address several questions. First, did moving fish under altered environments change their gene expression? For this we chose to move fish from the natural light conditions in the lake to the simplest light conditions of the fluorescent lights typical in labs (Natural to Narrow). Second, what aspect of the lab lighting affects the change in opsin expression (Narrow to Broad)? Is it the intensity of light or spectral content or both? We therefore chose to compare lights of the same spectra varying in intensity (dim (DIM) vs bright fluorescent illumination (BFLR)) and then varied spectral content (fluorescent vs metal-halide illumination, UV-rich).

Natural to Narrow

The natural light environment was estimated from measurements taken at Thumbi Island (14°1'8"S 34°48'53"E) [226]. *M. mbenji* are normally found in the top 20m of the lake, so side-welling irradiance measurements at 3m and 10m were used for comparison to the narrow spectrum lab environment [227]. The irradiance spectra for the lab fluorescent environment was measured from the underwater side-welling light in the tanks using a cosine collector attached to the Ocean Optics USB4000 spectrometer (Ocean Optics, Dunedin, FL, USA) with a 1 micron diameter, 2m UV transmissive fiber. The tanks were housed under Phillips 32W (Philips, Somerset, NJ, USA) fluorescent bulbs.

Narrow to Broad

Three light environments were chosen for this study- dim fluorescent (DIM), bright fluorescent (BFLR), and metal halide (UV). Similar to the Natural to Narrow study, the irradiance spectra were measured underwater for side-welling light. The DIM and BFLR treatment tanks were under Phillips 32W fluorescent bulbs and the UV treatment tanks were under broad-spectrum metal halide XM 10,000 K bulbs (XM lighting, Springfield, OH, USA). DIM tanks had an intensity of less than 0.2 klux as measured by a HI 97500 Portable Luxmeter (Hanna Instruments, Woonsocket, RI, USA). Intensity in BFLR and UV tanks was ~2.5 klux. The halide UV treatment was intended to simulate water conditions in the lake. It should be noted that the UV treatment was broadband illumination but still less intense compared to light incident on the Lake Malawi water surface and equivalent to intensity at a depth of 20m in the lake. All the light environments operated under a peak 12-12 day/night photoperiod.

4.3.3 Sample collection

Natural to Narrow

Adult *M. mbenji* from a single population were shipped to the University of Maryland after being held for several weeks in an outdoor greenhouse in Malawi, being kept in the dark during shipment. Immediately on arrival, 20 individuals were euthanized and phenotyped for opsin expression (DAY0). The remaining fish (n= 29) were placed in a single 50-gallon tank under fluorescent light conditions. The overall intensity of light was several orders of magnitude lower than that in the lake and importantly was UV deficient. 10 fish each were then sampled at 3 days (DAY3), 3 weeks (DAY21) and 5 weeks (DAY35) (the DAY35 sample had 9 individuals), to look for changes in opsin expression (Fig. 4.1A). The light conditions were kept constant for the entire duration of the experiment. To reduce time of day effects, all samples were collected between 3-5 PM. Individuals were sampled randomly with respect to sex since no correlation between sex and opsin expression has been found in wild-caught or lab-reared Lake Malawi cichlids . Specimens were euthanized using an overdose of MS222 and retinas were dissected and stored at -20 °C in RNAlater.

Narrow to Broad

Lab-raised *M. benetos* were F2 progeny of wild-caught fish that were more than 3 years old and resided in dim fluorescent tanks. Fish from 3 unrelated families with at least 2 individuals from each family were included in the baseline dim fluorescent (DIM) group. Specimens were either collected from DIM tanks or moved to bright fluorescent (BFLR) or metal halide (UV) tanks for 3 or 7 days prior to sampling (Fig. 4.1B). Samples were

collected using the same methods as Natural to Narrow with either 9 or 10 individuals collected for each treatment.

4.3.4 Quantification of opsin expression

Opsin expression was quantified using formerly described methods [158]. Briefly, RNA was extracted from dissected retinas using the QIAshredder and RNeasy Mini Kit (Qiagen, Valencia, CA, USA) and transcribed into cDNA (Superscript III, Invitrogen, Carlsbad, CA, USA). cDNA from each sample was quantified by real-time, quantitative PCR (RT-PCR) using Taqman primers (Life Tech, Carlsbad, CA, USA) and probes specific for each opsin on a Roche Lightcycler 480 (Roche, Bavaria, Germany). As in previous work, *RH2Aα* and *RH2Aβ* genes that are more than 99% identical were combined to become the 'RH2A' measurement. Plates were run with a mixture of experimental groups to reduce plate effects and each reaction was performed in duplicate. PCR efficiencies were determined by amplifying an opsin construct containing a single copy of each of the six opsins [43]. Expression of each opsin relative to single and double cones was calculated as a percent of the total opsin within that cell type. For example, the fraction, *f*, of *SWS1* expression in the single cones is calculated as:

$$f_{SWS1,\,SC} = T_{SWS1} / (T_{SWS1} + T_{SWS2B} + T_{SWS2A})$$

where T is the transcript level for each of the single cone opsin genes. Similarly, the fraction of *LWS* expression in double cones is:

$$f_{LWS,\,DC} = T_{LWS} / (T_{RH2A} + T_{RH2B} + T_{LWS})$$

To eliminate the effects of normalization to the opsin pool within a cone type, opsin expression was also normalized to another gene, *GNAT2* [55]. *GNAT2* is the cone-specific

alpha subunit of transducin, which is part of the photoreceptor signal transduction cascade and is therefore likely constantly and highly expressed in the cone cells of the retina. PCR efficiencies were calculated from RT-PCR measurements of serial dilutions of a mixture of *GNAT2* PCR product and the opsin construct. Using the PCR efficiencies (E_i) and critical numbers (Ct_i) for each opsin, the normalized cone opsin expression to *GNAT2* was calculated (in percentage) as:

$$T_i/T_{GNAT2} = (1+E_{GNAT2})^{Ct GNAT2} /(1+E_i)^{Cti}$$

where 'i' is the cone opsin of interest, E_{GNAT2} and Ct_{GNAT2} are the PCR efficiency and critical number for *GNAT2* respectively.

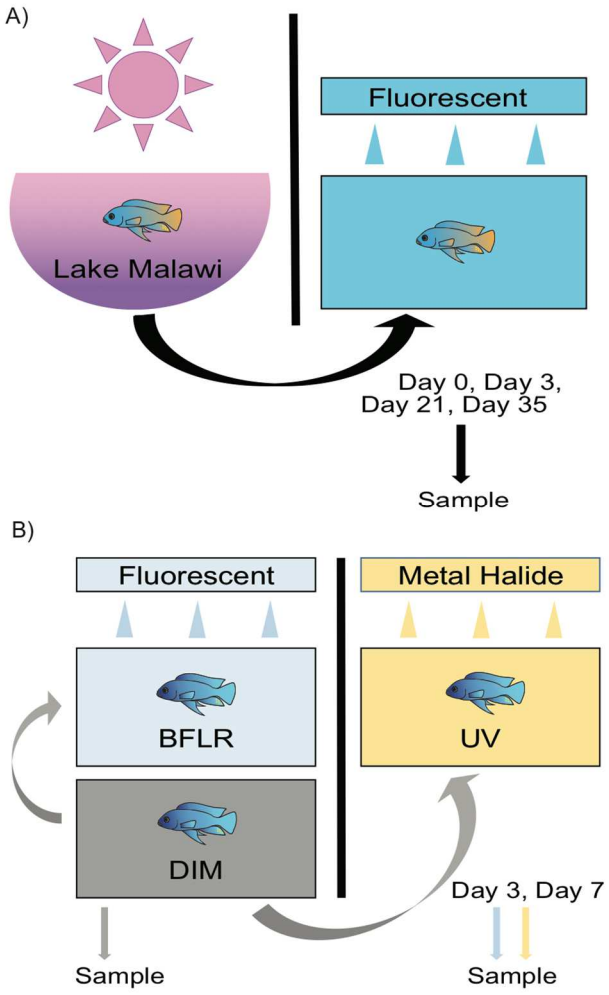


Fig. 4.1: Scheme for the experimental light conditions and sampling time points. A) Natural to Narrow- wild *M. mbenji* adults were collected from Lake Malawi and brought into the lab where they were sampled upon arrival (Day 0) and Days 3, 21 and 35. B) Narrow to Broad- lab-raised *M. benetos* adults residing in dim tanks (DIM) were either sampled for baseline or moved under bright fluorescent tanks (BFLR) or under metal halide lamps (UV). Samples were taken on Days 3 and 7.

4.3.5 Quantum catch calculations

Absolute quantum catch was estimated for each visual pigment (opsin protein bound to the A1 chromophore) to approximate how many incident photons each pigment would absorb in the different light environments [228]. Quantum catch thus helps measure how each visual pigment is stimulated in the wild and the different lab treatment conditions. Quantum catch is estimated using:

$$Q_i = \int I(\lambda)L(\lambda)R(\lambda)d\lambda$$

where I(λ) represents the normalized irradiance spectra of the environment, L(λ) is the lens transmission, and R(λ) represents the photoreceptor absorption values at each wavelength [229]. Since opsin gene sequences and visual pigment peak sensitivities are similar for different *Metriaclima* species, the λ_{max} of *Metriaclima zebra* visual pigments were used. For visual pigments not expressed in *M. zebra*, an average of sensitivities from different Lake Malawi species was used to calculate quantum catches [49]. Lens transmittance was measured from four lenses of adult lab-reared *M. benetos* [230]. Since *M. mbenji* occur in the top 20m of the water column all quantum calculations for Natural to Narrow experiment were performed at a depth of 3m and 10m.

4.3.6 Statistical tests

To test for significance of the change in opsin expression between each experimental light environment, post hoc Tukey's multiple comparisons of means with 95% family-wise confidence level was used. Statistical tests were performed in R version 3.2.3 (R Foundation for Statistical Computing, Vienna, Austria) and were corrected for multiple comparisons using the Bonferroni adjustment.

4.4 RESULTS

4.4.1 Changes in opsin expression

Overall, opsin expression varied as a function of both time and treatment. Significant differences in opsin expression were observed in both experiments. Results from post-hoc Tukey's multiple comparisons for all the treatment and time points are provided in Table 4.1 and 4.2.

Table 4.1: Post-hoc Tukey's comparisons for all the opsins for Natural to Narrow experiment. All the different treatments and time points are shown. Shown are the adjusted p-values for each opsin and each normalization procedure. All significant p-values (< 0.05) are bolded, and significant p-values after Bonferroni correction are denoted with an asterisk. For each opsin, the p-values are provided for both normalization procedures (total opsin within cell type and *GNAT2*).

A) Single cone opsins:

	SWS1$_{totSC}$	*SWS1*$_{GNAT2}$	*SWS2B*$_{totSC}$	*SWS2B*$_{GNAT2}$	*SWS2A*$_{totSC}$	*SWS2A*$_{GNAT2}$
DAY3 – DAY0	**0.0088**	**0.0002***	**0.0079***	0.9485	0.7643	0.5786
DAY21 – DAY0	0.0583	**0.0019***	0.0514	0.9518	0.5309	0.314
DAY35 – DAY0	**0.0187**	**0.0091**	**0.017**	0.5947	0.7694	0.6596
DAY21 – DAY3	0.9176	0.9136	0.9227	0.9999	0.9871	0.9783
DAY35 – DAY3	0.9989	0.7571	0.9989	0.9178	0.9999	0.9998
DAY35 – DAY21	0.9628	0.9849	0.9657	0.9141	0.9903	0.9668

B) Double cone opsins:

	RH2B$_{totDC}$	*RH2B*$_{GNAT2}$	*RH2A*$_{totDC}$	*RH2A*$_{GNAT2}$	*LWS*$_{totDC}$	*LWS*$_{GNAT2}$
DAY3 – DAY0	**0.0049***	**<0.0001***	0.2139	**0.0009***	0.5264	0.996
DAY21 – DAY0	0.8184	**0.0026***	0.603	**0.0107**	0.8444	0.3641
DAY35 – DAY0	0.0594	**0.0095**	0.1561	0.1343	0.9999	0.6941
DAY21 – DAY3	0.11	0.6888	0.9217	0.8942	0.2449	0.6101
DAY35 – DAY3	0.8874	0.5334	0.9968	0.4413	0.6233	0.8679
DAY35 – DAY21	0.4291	0.9922	0.8439	0.8443	0.9209	0.973

Natural to Narrow

When normalized to total opsins in single cones, *SWS2B* increased significantly within three days (Fig. 4.2C). While *SWS1* decreased within the same period, the change was not significant (Fig. 4.2A). While no significant changes in *SWS2B* were observed when normalized to *GNAT2*, expression of *SWS1* significantly decreased within three days (Fig. 4.2B), indicating that *SWS1* was downregulated. Following the three-day time period, *SWS1* and *SWS2B* remained unchanged for the duration of the experiment. No changes in *SWS2A* were observed, irrespective of the normalization (Fig. 4.3A, B).

When normalized to total opsin in double cones, *RH2B* decreased significantly after three days in the lab environment (Fig. 4.3C). Post three days, *RH2B* remains unchanged. When normalized to *GNAT2*, a significant decrease in expression was observed between DAY0 and both DAY3 as well as DAY21, denoting that *RH2B* was downregulated as well (Fig. 4.3D). *RH2A* showed drastically different patterns for each normalization procedure. When normalized to total opsin within double cones, *RH2A* showed no changes between arrival and the later time points (Fig. 4.3E). However, when normalized to *GNAT2*, a significant

decrease was observed between arrival and both DAY3 (Fig. 4.3F). Regardless of normalization, no changes in *LWS* were observed in the double cones (Fig. 4.3G, H).

Narrow to Broad

Adult fish show significant changes after treatment for only three days (Fig. 4.4). Only some opsins showed differences in expression in the different light environments. Longer duration treatments also seemed to increase the strength of differences.

In the single cones, *SWS1* increased in the UV environment. Both time points in the UV (UV3 and UV7) were significantly higher than DIM when normalized to opsin within single cones (Fig. 4.4A). With *GNAT2* normalization, UV7 had significantly higher *SWS1* than DIM treatment indicating that *SWS1* opsin was upregulated and that this upregulation increased with more time (Fig. 4.4B). When normalized to total opsin in single cones, *SWS2B* showed the opposite trend from *SWS1*, with expression in both UV3 and UV7 being significantly less than in DIM (Fig. 4.4C). However, no significant changes were observed in *SWS2B* when normalized to *GNAT2* (Fig. 4.4D). In *M. benetos*, *SWS2A* is only detected at very low levels in the retina. When normalized to opsin expression within single cones, expression in the BFLR7 was significantly higher than both the UV3 and UV7 (Fig. 4.5A). The *GNAT2* normalization showed a significant decrease in *SWS2A* in UV3 relative to DIM but the values are extremely small and likely not biologically significant (Fig. 4.5B).

When normalized to total opsin in double cones, *LWS* expression in the BFLR7 was significantly higher than both UV3 and UV7 (Fig. 4.4E). Compared to the baseline DIM,

LWS opsin showed a significant increase in the bright fluorescent environment but only for the 7-day treatment (BFLR7) suggesting that upregulation of *LWS* in the BFLR tanks is intensified with longer duration in the treatment (Fig. 4.4F). Neither *RH2B* nor *RH2A* showed consistent differences with treatment. Compared to the baseline DIM, when normalized to total opsin within double cones, both *RH2B* and *RH2A* remained unchanged in the UV treatment (Fig. 4.5C, E). While no changes were observed in *RH2A* expression with *GNAT2* normalization, *RH2B* decreased in BFLR7 and both UV3 and UV7 when compared to BFLR3 (Fig. 4.5D, F).

4.4.2 Quantum catches

The calculated quantum catch values suggest that visual pigments are stimulated differently in each light environment. The absolute quantum catch values of all treatment conditions for both the experiments are presented in Table 4.3.

Natural to Narrow

Although there is variation, in the wild, all visual pigments performed relatively well at both depths. Since light transmission decreases with depth, quantum catches at 3m were approximately three times higher for all opsins than those for 10m. Consistent with the higher irradiance levels towards the longer wavelength end of the spectra (Fig. 4.6A), quantum catches for double cone pigments (*RH2B*, *RH2A*β, *RH2A*α, *LWS*) were about 4 times higher than those for single cones (*SWS1*, *SWS2B*, *SWS2A*) (Fig. 4.6B). Owing to their red-shifted spectra, the difference in quantum catches doubled to about eight times between the two cone types under fluorescent lights (Fig. 4.6A). Except for *SWS1*, quantum

catch for all other opsins was ~1.2 orders of magnitude lower in the lab environment than in the wild. It is important to note that the quantum catch for *SWS1* falls to zero in the UV deficient lab environment (Table 4.3A). This means that in individuals shifted from the lake to the lab, the *SWS1* photoreceptors in the retina received no stimulation.

Narrow to Broad

The metal halide (UV) environment was relatively effective in simulating the environmental conditions in the wild. Although the intensity of the UV treatment (~ 2.5 klux) was at least 40 times lower than the shallow water conditions in the lake (~ 100 klux), the treatment was successful in providing a wider spectrum that stimulated all visual pigments (Fig. 4.7A). As expected, the fluorescent spectra for both the BFLR and DIM treatments were shifted heavily to longer wavelengths (Fig. 4.7A). Consequently, the quantum catches of the double cone pigments were higher than the single cone pigments (Fig. 4.7B).

Similar to *SWS1* in the Natural to Narrow experiment, there is zero quantum catch for *SWS1* in the DIM treatment (Fig. 4.7B, Table 4.3B). The quantum catch for *SWS2B* was lowest in the DIM, but nearly the same in the BFLR and UV treatments. For all other visual pigments, quantum catches were highest in the BFLR treatment (Fig. 4.7B). Of note, the quantum catch for *LWS* in BFLR was three times higher than in the UV treatment, and nearly nine times higher than DIM. Also, *SWS1* is the only visual pigment with quantum catch values higher in UV compared to all other treatments (Table 4.3B).

Table 4.2: Post-hoc Tukey's comparisons for all the opsins for Narrow to Broad experiment. All the different treatments and time points are shown. Shown are the adjusted p-values for each opsin and each normalization procedure. All significant p-values (< 0.05) are bolded, and significant p-values after Bonferroni correction are denoted with an asterisk. For each opsin, the p-values are provided for both normalization procedures (total opsin within cell type and *GNAT2*).

A) Single cone opsins:

	SWS1$_{totSC}$	*SWS1*$_{GNAT2}$	*SWS2B*$_{totSC}$	*SWS2B*$_{GNAT2}$	*SWS2A*$_{totSC}$	*SWS2A*$_{GNAT2}$
BFLR3 – DIM	0.8095	0.621	0.8298	0.9163	0.7961	0.9425
BFLR7 – DIM	0.0717	**0.0277**	0.0497	0.3117	0.148	0.5934
UV3 –DIM	**0.0018***	0.6364	**0.002***	**0.0085**	0.3738	**0.0073***
UV7 –DIM	**0.0009***	**0.0003***	**0.001***	**0.0429**	0.5647	**0.0183**
BFLR7 – BFLR3	0.5294	0.486	0.4124	0.8189	**0.0128**	0.9577
UV3 – BFLR3	**0.0444**	0.9999	**0.0437**	0.0841	0.9579	0.0608
UV7 – BFLR3	**0.0296**	**0.0278**	**0.0277**	0.2859	0.9967	0.1322
UV3– BFLR7	0.6697	0.4714	0.7787	0.5371	**0.0018***	0.2491
UV7 – BFLR7	0.593	0.6225	0.6997	0.8976	**0.0038***	0.4426
UV7 –UV3	0.9999	**0.0261**	0.9999	0.9562	0.9963	0.9928

B) Double cone opsins:

	RH2B$_{totDC}$	*RH2B*$_{GNAT2}$	*RH2A*$_{totDC}$	*RH2A*$_{GNAT2}$	*LWS*$_{totDC}$	*LWS*$_{GNAT2}$
BFLR3 – DIM	0.2133	**0.0422**	0.3249	0.1312	0.9999	0.5949
BFLR7 – DIM	**0.0173**	0.3479	0.9973	0.5904	**0.0108**	**0.0071***
UV3 – DIM	0.6019	0.8161	0.1609	0.9781	0.5229	0.8749
UV7 – DIM	0.1481	0.5442	**0.0139**	0.1667	0.2992	0.8259
BFLR7 – BFLR3	**<0.0001***	**0.0002***	0.2036	0.8851	**0.02**	0.2451
UV3 – BFLR3	**0.0089**	**0.003***	**0.0015***	0.3905	0.4601	0.1502
UV7 – BFLR3	**0.0006***	**0.0006***	**<0.0001***	0.9997	0.256	0.1126
UV3 – BFLR7	0.399	0.9371	0.3163	0.9071	**0.0001***	**0.0006***
UV7 – BFLR7	0.8631	0.9956	**0.0409**	0.9381	**<0.0001***	**0.0003***
UV7 – UV3	0.9145	0.9934	0.8758	0.4694	0.9969	0.9999

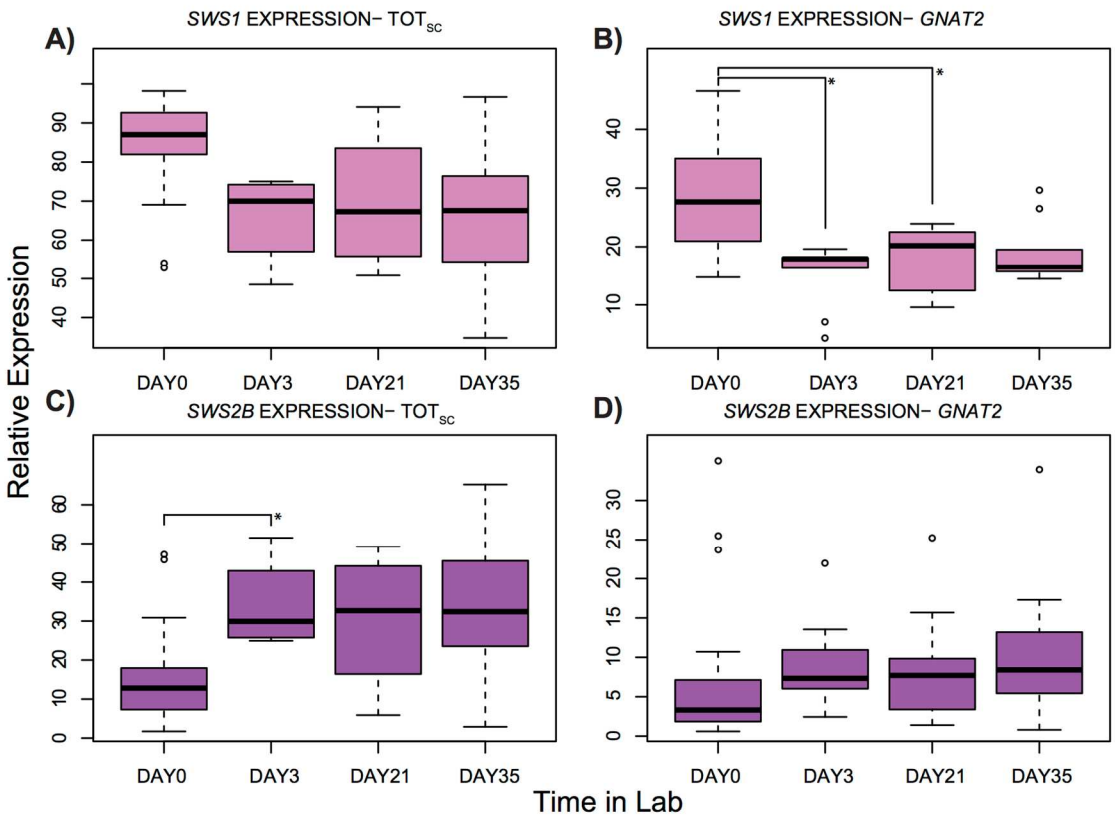


Fig. 4.2: Natural to Narrow opsin expression for *SWS1* and *SWS2B* opsins. A) *SWS1* and C) *SWS2B* relative to total single cone expression, B) *SWS1* and D) *SWS2B* relative to *GNAT2* expression. Significant Tukey's multiple comparisons after Bonferroni corrections are indicated with an asterisk.

4.5 DISCUSSION

4.5.1 Plasticity in opsin expression in adult cichlids

Using quantitative PCR, we have shown that adult cichlids from Lake Malawi exhibit plastic changes in opsin expression mediated by differences in spectral irradiance. Even more striking is that these plastic changes occurred relatively fast, with some changes recorded three days after being placed in the new environment. This suggests that retinal sensitivities can change on a rapid timescale, enabling adult fish to adjust to short-term changes in lighting conditions. This might be helpful for fish swimming to different depths

to find new habitats or from changes in local water clarity from runoff during seasonal rains or algal blooms.

Table 4.3: Quantum catches of the opsins for different lighting environments. Quantum catches are provided in photons/cm^2/msec.

3.3A: Natural to Narrow

Opsin	Lake Malawi @ 3m	Lake Malawi @10m	Lab environment
SWS1	6.36E+11	1.78E+11	0
SWS2B	3.24E+12	9.39E+11	5.39E+10
SWS2A	5.41E+12	1.71E+12	1.02E+11
RH2B	9.94E+12	3.25E+12	2.13E+11
*RH2Aβ**	1.2E+13	3.85E+12	3.49E+11
*RH2Aα**	1.22E+13	3.88E+12	3.84E+11
LWS	1.2E+13	3.45E+12	4.99E+11

3.3B: Narrow to Broad

Opsin	DIM	BFLR	UV
SWS1	0	1.04E+10	2.37E+10
SWS2B	3.83E+09	1.5E+11	1.65E+11
SWS2A	1.34E+10	2.6E+11	2.06E+11
RH2B	4.32E+10	5.47E+11	3.3E+11
*RH2Aβ**	9.79E+10	9.18E+11	4.02E+11
*RH2Aα**	1.13E+11	1.03E+12	4.26E+11
LWS	1.68E+11	1.45E+12	5.28E+11

* Although quantum catches are calculated separately for *RH2Aα* and *RH2Aβ* opsins, the opsins were combined in the RT-PCR expression assay.

As mentioned earlier, most of our previous studies have measured opsin expression relative to the total pool of opsins. The reasoning for this is that color vision is contingent on the differences in the inputs received by the cone cells in the retina [228]. Differences in the expression of each cone type relative to total opsin either echo differences in the relative abundance of specific cone cells in the retina or more likely the packing of the opsin pigment in the outer segments of cones, which could change color perception. The downside of this normalization is that it obfuscates the mechanism by which the differences in visual tuning through opsin expression arise. Since expression changes of each opsin are no longer independent, it becomes impossible to predict which opsin is being up/down

regulated. To circumvent this problem, we have normalized opsin expression to the expression of a cone specific gene, *GNAT2*, similar to what has been done previously [225]. *GNAT2*, as part of the phototransduction cascade, is constantly expressed at high levels in the cone cells of the retina and should be expressed in conjunction with the cone opsins thus varying minimally between samples and between treatments. Indeed, results from 27 cichlid retinal transcriptomes confirm that *GNAT2* is relatively constant in expression (data not shown). It has a coefficient of variation comparable to that of beta-actin (and better than several other genes often used as housekeeping genes including *EF1a*, *PAX6*, and *UBE2z*). In the Natural to Narrow experiment, although normalization to total single cone opsins indicates that *SWS2B* increases significantly after just three days under fluorescent lights, *GNAT2* normalization suggests that adult animals only decrease *SWS1* while keeping *SWS2B* constant.

Plastic changes in opsin expression could be explained by several mechanisms. It is possible that the outer segments of the photoreceptors could increase in size. Such changes in photoreceptor length have been previously observed following long-term exposure to highly intense monochromatic light [221]. This is unlikely within the time frame observed within our study (three days). Another option is that the expression differences arise from changes in the relative numbers of cone types, with additional cone cells being added to the retina. Again, this too is highly improbable, as it has been demonstrated that complete differentiation of cone cells in gold fish takes approximately three weeks [231]. A third possibility is an increase in the density of the visual pigment in the cone photoreceptors. Since optimal opsin protein packing in the photoreceptor cilia is critical for proper

functionality of the phototransduction cascade and kinetics of light response, this seems unlikely as well [232]. We hypothesize that the plastic differences in expression arise from the expression of multiple opsins in the same photoreceptor, i.e. co-expression [103]. Earlier studies in the lab have shown that *M. zebra* coexpressed opsins in their cone cells and that this level of coexpression increased with altered environmental background spectra [55]. Since *M. benetos* and *M. mbenji* are from the same genus, it seems likely that plastic changes could cause an increase in coexpression of opsins in these individuals as well. Future studies with in-situ hybridization or electrophysiology from individual cone photoreceptors might throw some light on the cellular modifications underlying plasticity.

In this study, it is clear from the quantum catches that the largest difference between the two experiments arises from the absence/presence of ultraviolet illumination. As expected, major plastic changes are observed in the single cones, principally along the *SWS1-SWS2B* axis (ignoring the very low levels of *SWS2A* expression that are not likely to be physiologically relevant). Although quantum catches of the double cone opsins vary with light treatments, the magnitude of change is much smaller. Further, this problem is likely compounded by the fact that the two *RH2A* opsins have been merged in our study. Future experiments with more severe shifts in the spectral space bounded by the appropriate wavelengths of the double cones might lead to more pronounced plasticity.

Both intensity and spectral composition of the incident light seem to play a role in mediating plastic changes, with longer duration times generally enhancing the differences in opsin expression. Previous studies have speculated that the spectral content of light

rather than intensity played a bigger role in plasticity [53]. In the current study, we showed that at least for some opsins, namely *LWS*, intensity of light might also play a part (fish in the BFLR7 treatment have high *LWS* expression compared to the DIM). Additional plasticity studies on adult species are underway to help further delineate the role of spectral makeup and intensity in stimulating expression plasticity.

Lastly, while the light changes dramatically between Lake Malawi and the lab, there are several other factors that differ between those two environments including temperature, salinity, circadian period, diet, and presence or absence of predators. Although unlikely, there is a possibility that any of these factors could have an effect on opsin expression in the Natural to Narrow experiment. Given that complementary changes in opsin expression were seen in the Narrow to Broad experiment where all the lab conditions were the same and only the light composition was changed, we postulate that changes in opsin expression in Natural to Narrow were most likely due to the light environment differences.

4.5.2 Genetic inferences of plasticity

From this study, it is evident that *SWS2A* is never turned on in *Metriaclima* species in spite of the further improvements it would make to the quantum catch efficiency. This result is not novel and has been formerly observed when broods from other rock-dwelling cichlids were raised under fluorescent lights [53]. Also, this is consistent with the developmental program in *Metriaclima* that entails an ontogenetic progression from *SWS1* in larvae to *SWS1-SWS2B* as adults in the single cones but never to *SWS2A* expression [51]. These findings suggest that in Lake Malawi cichlid species with short wavelength sensitivities,

SWS2A is controlled by an independent set of factors that are different from the genetic factors that co-ordinate both *SWS1* and *SWS2B*. In fact, this hypothesis is supported by data from a recent hybrid cross in the lab between two genera that express *SWS1* and *SWS2B* [66]. Since *SWS1* and *SWS2B* are expressed in the same cone cell, it makes sense that they would share some, if not all, genetic factors.

Furthermore, plasticity experiments in killifish have illustrated that the extent of plastic opsin expression varies among families proposing that plasticity is under genetic control [212]. Though traditionally debated, it is now accepted that plasticity loci do exist and studies examining the genetic basis of plasticity have become common [198, 233, 234]. At present, we have a hybrid cross between a plastic species *M. mbenji* and a non-plastic species *A. baenschii* [66]. Splitting the F2 broods of this cross and raising them under UV-rich and UV-deficient lighting will help map genetic factors contributing to the plasticity of *SWS1/SWS2B* expression.

4.5.3 Comparing developmental and adult plasticity

A summary of what we currently know about developmental and adult plasticity is shown in Fig. 26. The results from this current study are compared against previously collected developmental progression in the lab and wild adult data in the context of *SWS1* expression [51, 53, 123, 235]. A comparison of the Natural to Narrow results to the *M. benetos* lab developmental series can be informative about how plastic changes in opsin expression are mediated during adulthood versus during development (Fig. 26 left) Lab-raised *M. benetos* larvae (10- 20 days) show the highest levels of *SWS1* (~95%, *M. benetos* lab development

in Fig. 26). As the larvae grow older in the lab, the levels of *SWS1* steadily decrease with adults (150-180 days) having an average *SWS1* closer to 50% (similar to *SWS1* in lab-raised *M. mbenji* adults, Fig. 26). Currently, no wild developmental data is available for comparison, but wild caught adult *Metriaclima* species have an average *SWS1* of around 85% (wild Thumbi, wild Mazinzi and wild Mbenji adults in Fig. 26). Since wild-caught adult *Metriaclima* species have high *SWS1* as adults, it is probable that they retain a predominance of *SWS1* throughout development and into adulthood.

It is therefore likely that the *M. benetos* lab developmental series shows the effects of plasticity in opsin expression during development. This is equivalent to the Natural to Narrow experiment in the current study that shows plasticity in opsin expression as adults. A closer look at these two parallel studies yields some interesting preliminary trends. The temporal dynamics of developmental plasticity seem to be slower than that of adult plasticity. One possible explanation for this is that during development, the plastic response is occurring in opposition to a preprogrammed genetic program, resulting in slower change. In adults, no such opposition exists, resulting in a quicker change in *SWS1* in the absence of the required UV signal. More work on the genetic basis and gene regulatory networks of opsin expression and plasticity might provide further insights.

In our Narrow to Broad experiment, *SWS1* increased in adult lab-raised fish to 58% (BFLR) and 67% (UV) in seven days (Fig 26 right). Additional work would establish whether lab-raised adults can ever achieve the wild expression levels (~85%) with increased light intensity. In the current study, the UV environment only emulated Lake Malawi quantum

catches at a depth of 20m, so it is possible that providing a stronger source of illumination that offers improved quantum catches for *SWS1* might produce a stronger plastic response.

4.5.4 Ecological context of plasticity

In the present study, we have shown that changes in the light environment leads to plasticity, which suggests some mechanism by which light is sensed and integrated into signals that both promote and inhibit opsin gene expression. This mechanism seems to be in place during both development and adulthood, allowing cichlids to be responsive to the environment throughout their lives and not just during a narrow sensitive developmental window. On short ecological time scales, such as one lifetime, opsin expression plasticity could be important for fish experiencing seasonal or man-made shifts in water quality that drastically change the light environment [219, 236]. Such seasonal changes in water quality and light attenuation in Lake Malawi have been previously reported [237]. Whether the seasonal effects on the light spectrum in Lake Malawi actually cause shifts in opsin expression in cichlids needs to be studied in more detail. Nonetheless, analogous findings in both the lab and Lake Victoria have shown that the spectral transmittance of water plays a major role in the interpretation of color signals in cichlids, a key element in mate choice and recognition [62, 238].

Additionally, the depth of Lake Malawi also provides opportunity for shifts in light environment within species ranges. *Metriaclima* species typically inhabit the top 20m of the lake, which shows ever-decreasing amounts of ultraviolet light and therefore less efficiency of the *SWS1* photoreceptor to perceive light with depth [226]. A fish moving

down to inhabit a different part of the habitat range could be enough to induce plasticity similar to what we have demonstrated in this study. Wild *Metriaclima* species in the lake typically express miniscule amounts of *SWS2B* and large amounts of *SWS1* [53]. A fish moving to a habitat niche that opens up around 15-20 meters in the clear waters of Lake Malawi would likely have very little ultraviolet light. This would necessitate a speedy shift in expression from *SWS1* to *SWS2B* in the single cones, similar to what we observed in the Natural to Narrow experiment. Interestingly, Zimbawe rock is one location in Lake Malawi where adult *M. zebra* were found to have only 55-60% *SWS1* in their retinas [235]. This pinnacle is very steep and extends to some of the deeper *M. zebra* habitats in the lake, supporting that depth could induce shifts from *SWS1* to *SWS2B*.

4.5.5 Evolutionary implications of plasticity

Cichlids are famous for their incredibly accelerated adaptive radiation in the African Rift Lakes [146, 147]. Along with opsin expression as depicted in the present study, plasticity acts across a broad range of cichlid phenotypes. These include trophic morphology [239], body shape [240], gill and brain size in response to hypoxia [241], foraging [233] and social behavior [242]. Plasticity allows an organism to quickly respond to an environmental shift instead of changing a phenotype through purely genetic means (mutation in the coding or regulatory sequences), which would simply be too slow. Over time, selection can then act on the range of phenotypes produced by plasticity to canalize them by genetic assimilation or allow them to remain plastic [2, 4, 243, 244].

Ancestral state reconstruction of the African cichlid tree reveals that the ancestral riverine species that seeded Lake Malawi likely possessed the long opsin combination (*SWS2A*, *RH2B*, *LWS*) [58]. The Nile tilapia is largely considered to be an extant representative of this riverine ancestor. Despite having the long opsin combination in adulthood, Nile Tilapia expresses all seven opsins at some point in its lifetime [51]. Additionally, Tilapia seems to exhibit developmental plasticity, but has lost adult plasticity in opsin expression [245]. It is likely that the developmental plasticity possessed by this riverine ancestor allowed canalization of the array of potential opsin expression phenotypes into the different opsin combinations seen in Lake Malawi today. For example, *Metriaclima* species typically express the short combination as adults (*SWS1*, *RH2B*, *RH2A*), which is the same as larval expression observed in Nile tilapia [51]. Perhaps *Metriaclima* have genetically solidified the ancestral neotenic expression pattern for use in adults. Neoteny may also explain the mechanism of how *Metriaclima* could have maintained plasticity in adulthood.

Retaining plasticity is likely costly [246, 247] so why preserve the ability to change opsin expression into adulthood? Moreover, if not utilized, plasticity is likely to be lost through drift [243]. Selection to maintain plasticity has to overcome these prohibitive costs or suffer random loss. The geological history of Lake Malawi provides some context for the possible selective force to sustain plasticity. In the last 1.3 million years lake levels have continuously changed; at times dropping 600m to confine the lake to one 100m deep pool, and at other times filling to 200m above current levels [248]. Further characterization has shown dry periods of "green lake" conditions with low lake levels and high eutrophication punctuated by about 100,000 year intervals of the deeper "blue lake" conditions much like

the lake of today [249]. With the extreme shifts in light environment that would occur during these alternating blue and green lake periods, it is not hard to speculate that the ability to switch visual sensitivity through plasticity would be important. Theory predicts that stable environments will favor genetic assimilation while plasticity is more likely to be maintained in unstable environments [243, 247]. It seems likely then that Lake Malawi would be a good candidate for retention of plasticity in different cichlid lineages.

4.6 CONCLUSION

Light-induced plasticity seems to be a powerful fine-tuning mechanism that allows cichlids to visually evaluate their immediate surroundings more efficiently. Since vision is critical for cichlid fitness, this work brings attention to the connection between plasticity and rapid diversification. Lastly, studying phenotypic change in response to an environmental shift in the context of adaptive radiations goes a long way into consolidating plasticity into modern evolutionary biology.

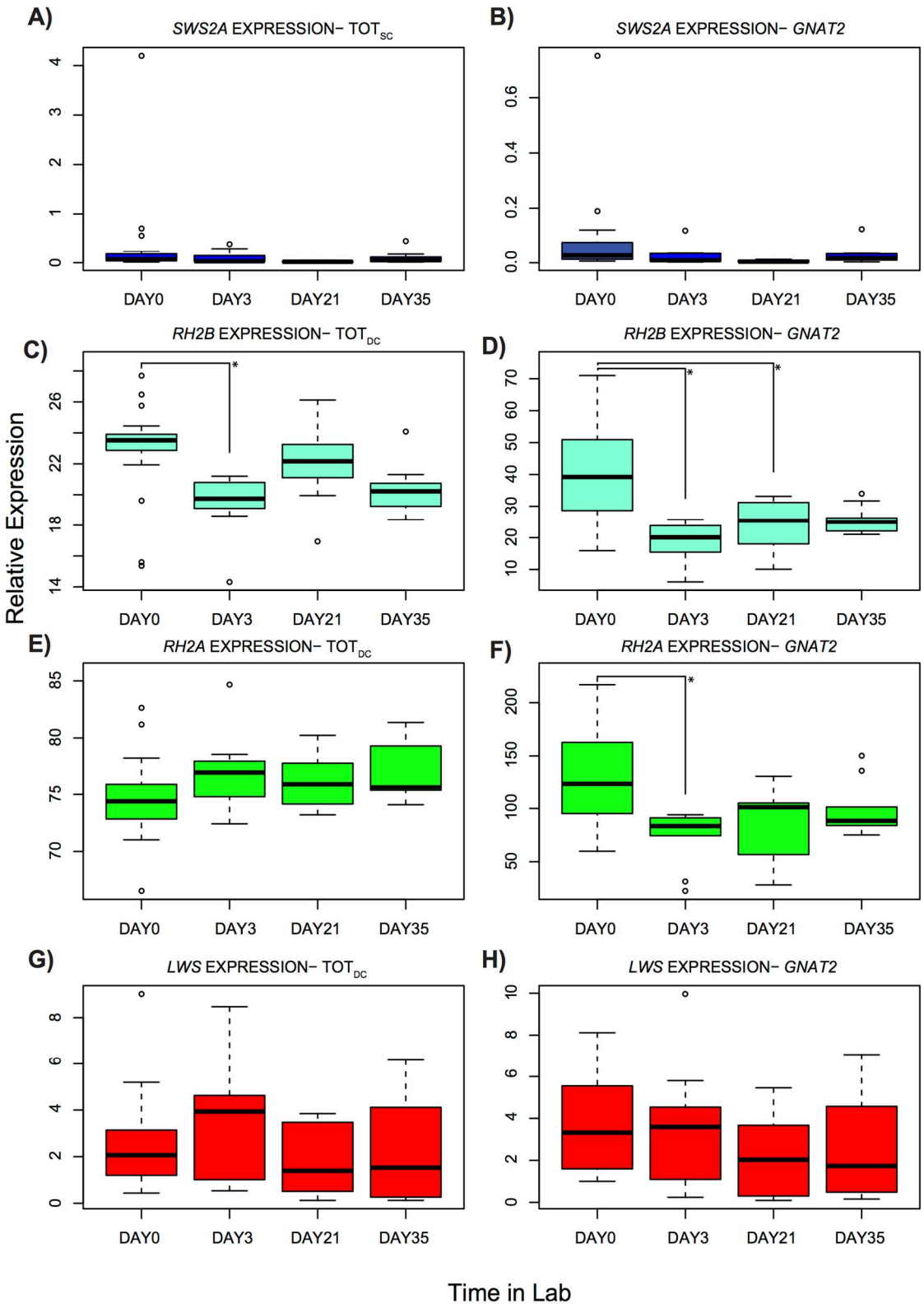

A)
SWS2A EXPRESSION– TOT$_{SC}$
B)
SWS2A EXPRESSION– GNAT2
C)
RH2B EXPRESSION– TOT$_{DC}$
D)
RH2B EXPRESSION– GNAT2
E)
RH2A EXPRESSION– TOT$_{DC}$
F)
RH2A EXPRESSION– GNAT2
G)
LWS EXPRESSION– TOT$_{DC}$
H)
LWS EXPRESSION– GNAT2
Relative Expression
DAY0
DAY3
DAY21
DAY35
Time in Lab

Fig. 4.3: Natural to Narrow opsin expression for *SWS2A, RH2B, RH2A* and *LWS* opsins. A) *SWS2A* relative to total single cone expression; C) *RH2B,* E) *RH2A* and G) *LWS* relative to total double cone expression. B) *SWS2A,* D) *RH2B,* F) *RH2A,* and H) *LWS* relative to *GNAT2* expression. Statistics are Tukey's multiple comparison of means with *p<0.05 and **p<0.01.

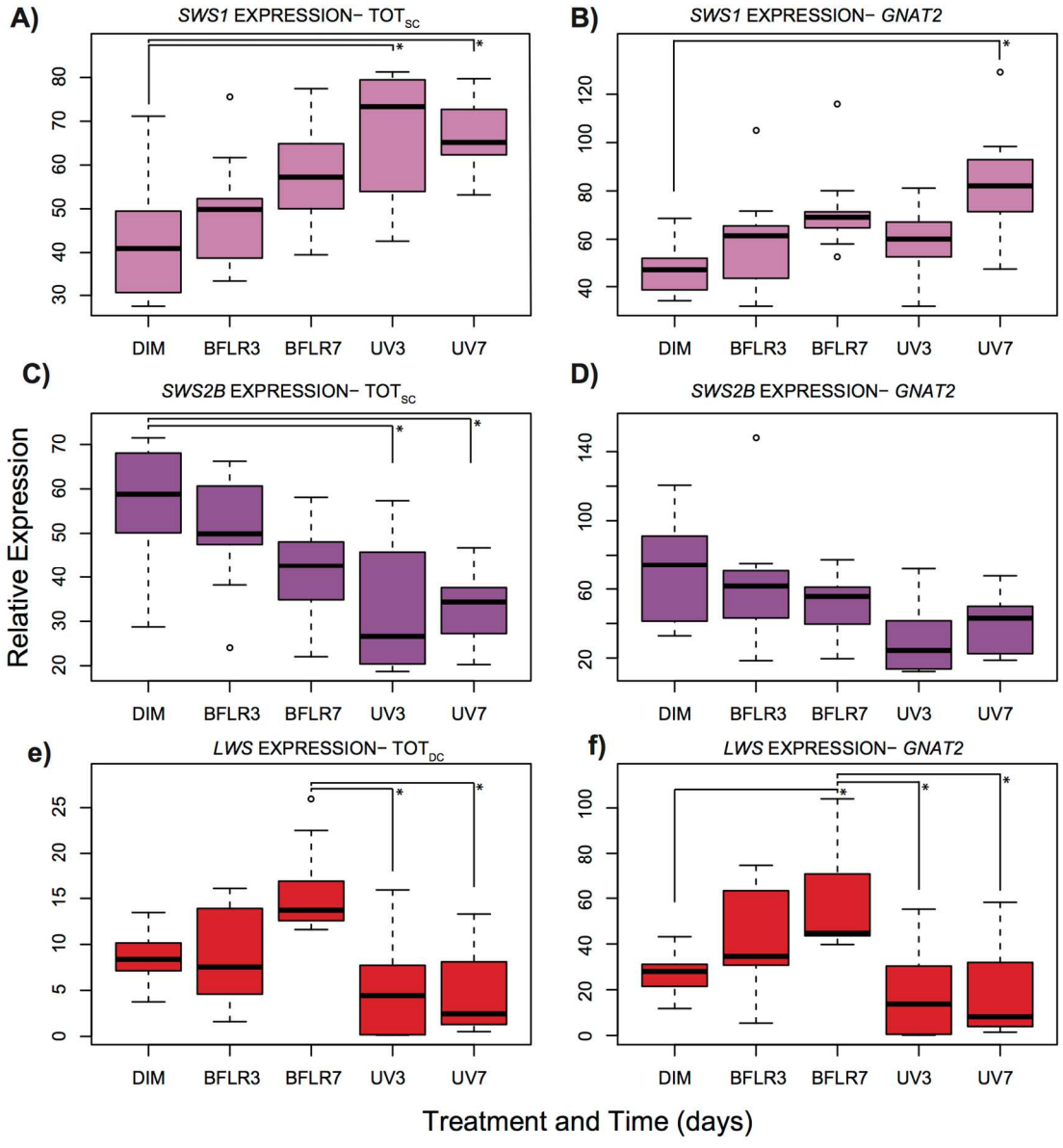


Fig. 4.4: Narrow to Broad opsin expression for *SWS1*, *SWS2B* and *LWS* opsins. A) *SWS1* C) *SWS2B* and E) *LWS* relative to total double cone expression, as compared to B) *SWS1* D) *SWS2B* and F) *LWS* relative to *GNAT2* expression. Significant Tukey's multiple comparisons after Bonferroni corrections are indicated with an asterisk. BFLR3/7- Bright fluorescent environment Day 3/Day7, UV3/7 – Metal halide UV-rich environment Day 3/Day 7.

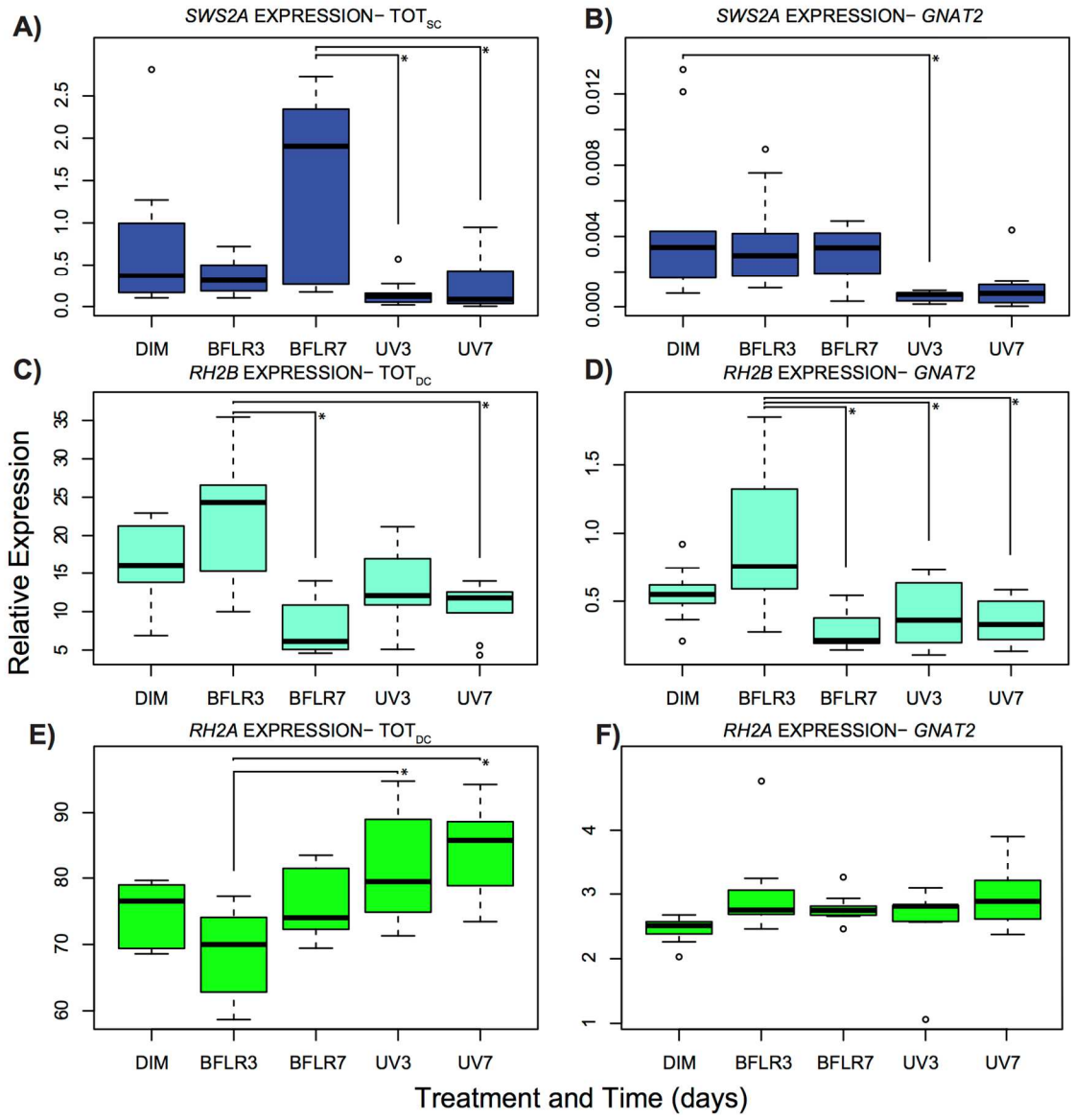


Fig. 4.5: Narrow to Broad opsin expression for *SWS2A*, *RH2B* and *RH2A* opsins. A) *SWS2A* relative to total single cone expression, C) *RH2B* and e) *RH2A* relative to total double cone expression. B) *SWS2A,* D) *RH2B,* and F) *RH2A* relative to *GNAT2* expression. Statistics are post-hoc Tukey's multiple comparison of means with significance indicated *p<0.05 and **p<0.01.

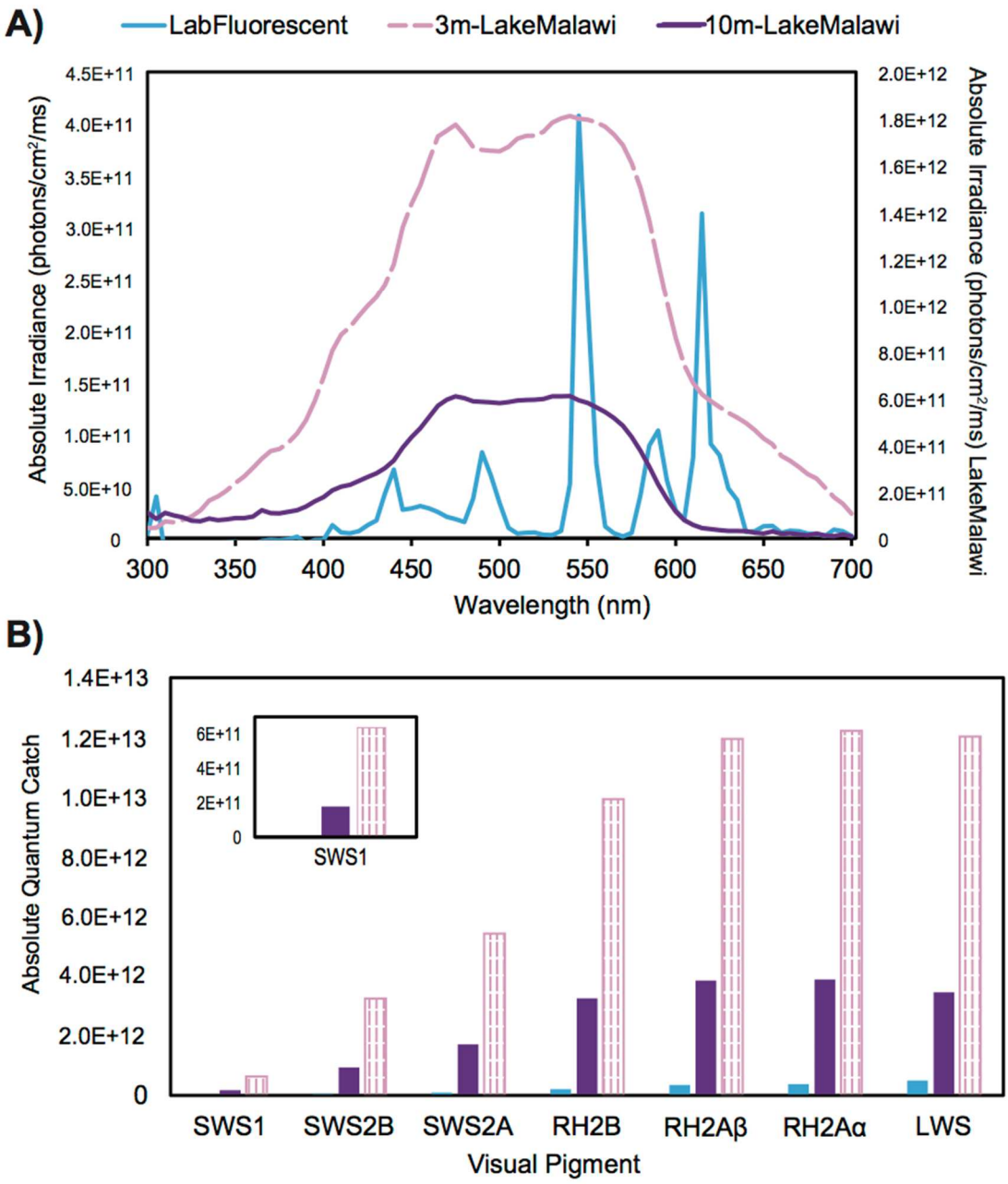


Fig. 4.6: Natural to Narrow light environment spectra and quantum catch. A) Lake Malawi spectra are measured on the secondary axis. Side-welling irradiance is provided in photons/cm^2/msec. Light environment spectra were used to calculate absolute quantum catch for each visual pigment. B) The lab fluorescent environment is greatly diminished in quantum catch for every visual pigment compared to Lake Malawi. Inset shows the close-up of the *SWS1* quantum catch.

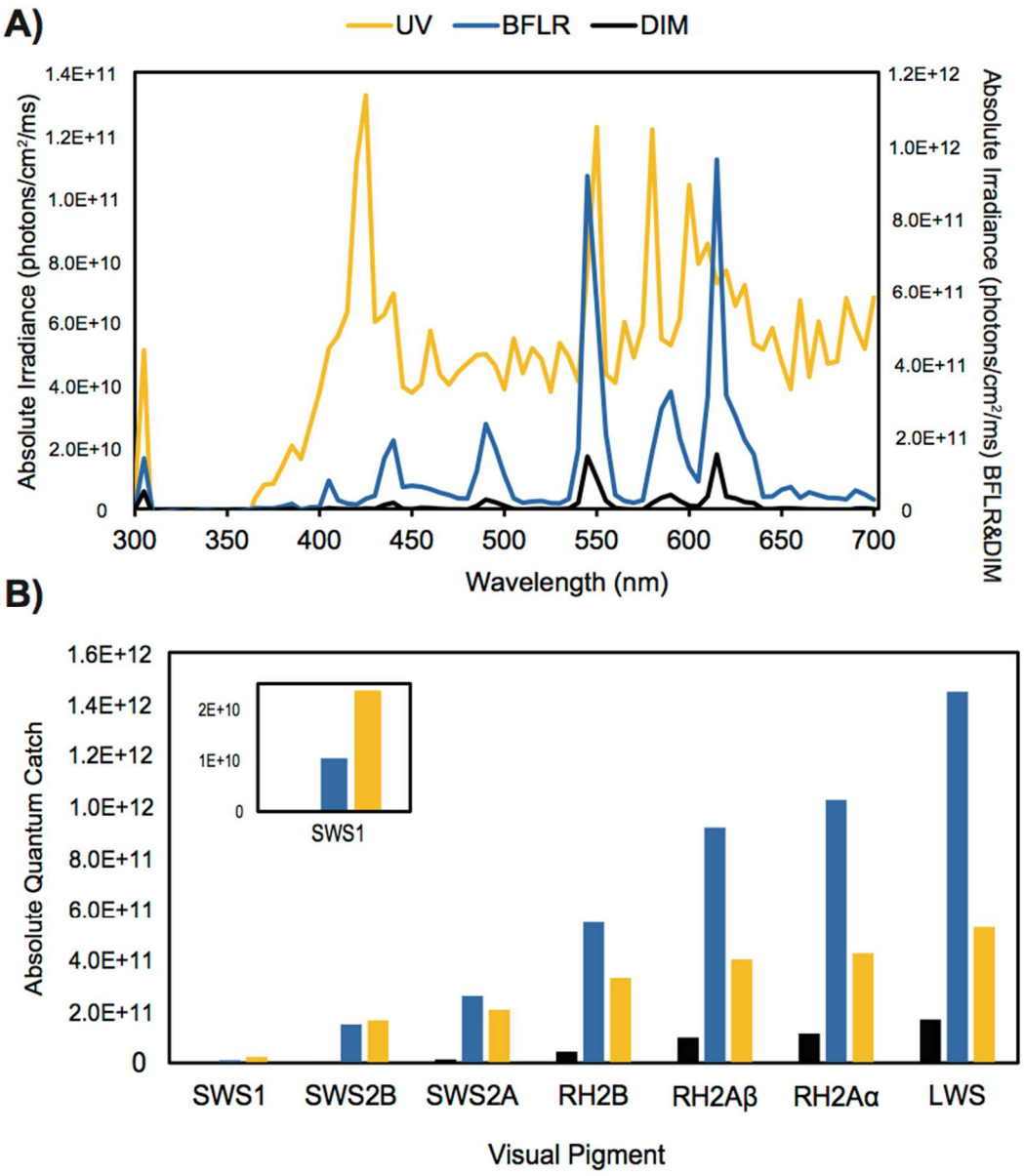


Fig. 4.7: Narrow to Broad light environment spectra and quantum catch. The spectra of the UV environment is much broader, but less intense than BFLR. A) BFLR and DIM environment spectra are measured on the secondary axis. Side-welling irradiance is provided in photons/cm^2/msec. Light environment spectra were used to calculate absolute quantum catch for each visual pigment. B) BFLR had the highest quantum catch for each visual pigment except for *SWS1* where the UV environment had the highest. Inset shows the close-up of the *SWS1* quantum catch.

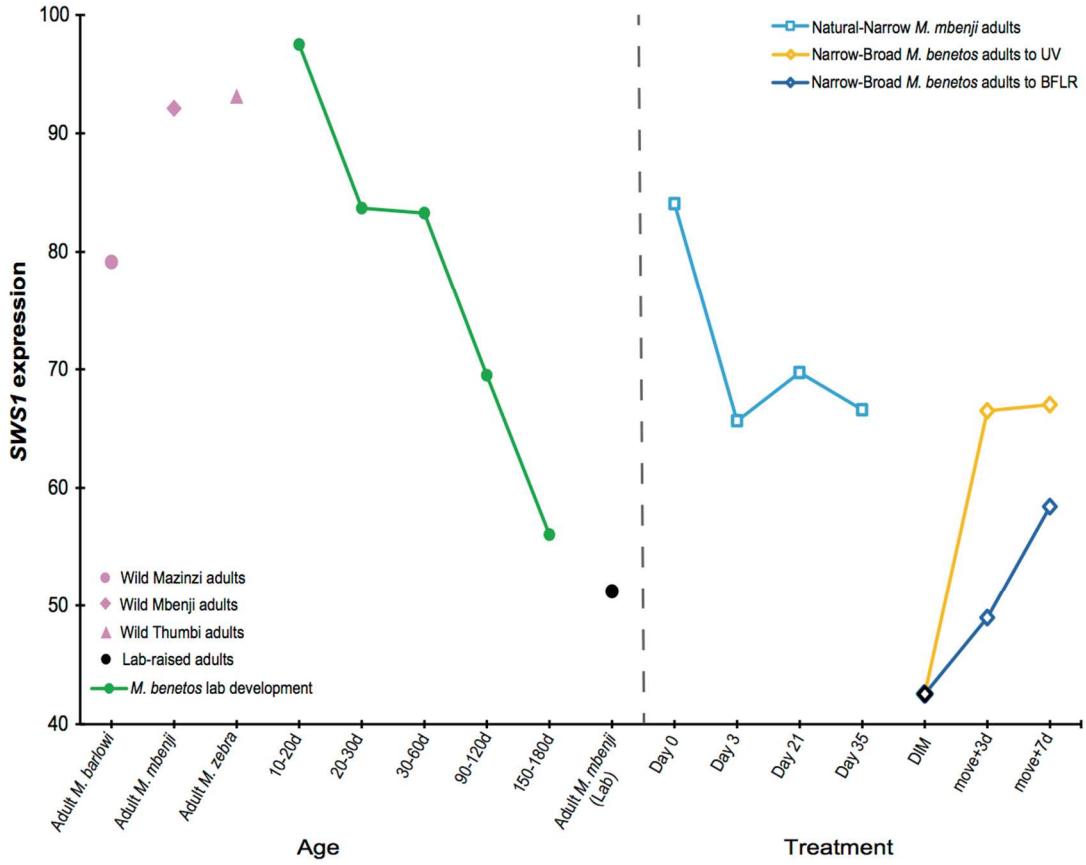


Fig. 4.8: A summary of current knowledge about Lake Malawi *SWS1* opsin expression in the context of development and plasticity. Open squares represent findings from the present Natural to Narrow and Narrow to Broad studies (right of dashed line) compared to wild adults, lab adults, and lab developmental series (left of dashed line). Wild adults include *M. mbenji*, *M. zebra*, and *M. barlowi*. We do not have wild caught *M. benetos* for comparison. The lab developmental series for *M. benetos* is similar to that of other *Metriaclima* species.

Funding

This work was supported by the National Eye Institute at the National Institutes of Health (NIH R01EY024639 to K.L.C).

Author's Contributions

S.P.N designed the Natural to Narrow experiment, and M.R.Y designed the Narrow to Broad experiment. Both S P.N and M.R.Y performed analyses and drafted the manuscript. K.L.C conceptualized the study and critically reviewed the article. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Thomas Kocher, Richard Zatha, Brian Dalton and Justin Marshall for help in collecting the fish and measuring light levels from Lake Malawi. We would extend our thanks to The Stuart M Grant Fish Farm for sending us wild caught fish from the lake. We would like to thank Daniel Escobar Camacho and Benjamin Sandkam for valuable suggestions on the manuscript. We thank Justin Marshall, Fabio Cortesi, Sara Stieb, and Martin Luehrmann for discussions and encouragement to further examine opsin plasticity.

<u>Chapter 5</u>: A second gene contributing to the differential expression of *SWS2A* expression in Lake Malawi cichlids.

Nandamuri SP, Schulte JE, Sandkam BA, Conte MA and Carleton KL (In progress)

5.1 ABSTRACT

African cichlids provide an ideal platform for studying the genes responsible for species divergence. These highly speciose fishes exhibit substantial variation in the expression of seven distinct cone opsin genes. Each species expresses a subset of the seven genes leading to extensive diversity in their visual sensitivities. In a genetic cross between two Lake Malawi cichlid species with varying visual palettes, we previously identified two quantitative trait loci (QTLs) responsible for the differential expression of the *SWS2A* opsin gene. Using fine-mapping and RNA-seq in retinal transcriptomes, we identified that *RX1* gene on LG23 is the causative gene for the first QTL. The present study focuses on the second QTL on LG5 with the aim of finding the second causative factor. Following fine-mapping, we identify three potential candidate genes based on differential expression in the parents of the cross and their functionality in photoreceptors. We present evidence for the likelihood of each candidate being responsible for the divergence in *SWS2A* expression.

5.2 BACKGROUND

A major challenge in evolutionary biology is comprehending how naturally occurring variation in DNA translates to phenotypic diversity. Most of the studies that try to connect genotype to phenotype for complex traits rely on linkage or association studies and mainly focus on estimating the effects at single loci. However, every gene works as part of a larger gene network and phenotypic variation could result, at least in part from a perturbation of highly dynamic and interconnected nodes within this network. Understanding how genes within this network are located, connected and work together to modulate the phenotype is critical step towards dissecting the genetic architecture of complex traits [250]. This information is crucial towards understanding how the trait responds to evolutionary forces such as drift/natural selection.

Sensory systems, such as vision serves as an interface between an organism and its environment, relaying useful information important for its survival and well-being. Animals have to adapt to their photic environments to succeed at a variety of fitness-related tasks such as foraging, communication, predator evasion and mate selection [24, 25, 251]. Visual pigments in the photoreceptors facilitate the beginning of visual perception, being composed of a transmembrane protein, the opsin and a vitamin-A derived chromophore [20]. Isomerization of the chromophore by light quanta leads to the activation of the opsin molecule, resulting in a visual transduction cascade that is ultimately perceived as neural output. Extensive work in model organisms has finely elucidated the genetic mechanisms by which multipotent precursor cells differentiate into retinal cells/subtypes [86, 87, 252, 253]. Particularly, molecular details of gene regulatory networks underlying photoreceptor

development have been well established [254-256]. These studies show that the photoreceptor differentiation network has a highly intricate and complex architecture. However, most of these network studies have been carried out in mice that have rod-dominated retinas. Consequently, genes underlying the development of cone photoreceptors remain under-studied. Additionally, most studies using lab models are limited to in-bred lab-reared populations and thus do not present an accurate picture of the genetic variants segregating in natural populations.

For this work, we aim to elucidate the gene networks underlying cone photoreceptors in African cichlids, which are a group of highly speciose freshwater fishes [79]. Incredibly expedited parallel radiations of cichlids, comprising several thousands of species, have arisen in the three main African rift lakes- Malawi, Tanganyika and Victoria in the last few million years [81, 146, 147]. The recent divergence of cichlids makes it possible to generate viable crosses among species and even between genera [50, 84, 124]. Due to their recent evolutionary history, key polymorphisms are likely to be shared between species [80]. This makes it possible to find underlying causal mutations by association mapping.

African cichlids have some of the most variable visual sensitivities among vertebrates [49, 50, 85, 257, 258]. All cichlids have seven cone opsin genes that are representative from the four vertebrate cone classes (SWS1, SWS2, RH2 and LWS) [43]. Individual species express a subset of the seven cone opsin genes in their retinas, leading to huge differences in the visual sensitivities between species [49]. This variation in opsin expression is

adaptive and aids visual communication for critical tasks such as finding food and potential mates [59, 62].

To identify genes responsible for regulating the differential expression of *SWS2A* opsin, we focus on a previously reported genetic cross between two sand-dwelling Lake Malawi cichlid species - *Tramitochromis intermedius* and *Aulonocara baenschi*, (TA cross) [50]. *A. baenschi* and *T. intermedius* have different opsin expression profiles, with species expressing *SWS2B* or *SWS2A* opsins in their single cones respectively. A genome-wide quantitative trait locus (QTL) study used restriction-associated DNA sequencing (RAD-seq) on the F2 hybrids of the TA cross to examine the differential expression of *SWS2A* cone opsin [63]. This yielded two QTL- one on LG23 and the other on LG5. The QTL on LG23 is the larger (LOD= 25.7) of the two QTL and explains nearly 30% of the variation in *SWS2A* expression among the F2 progeny. Recently published results suggest that *Retinal anterior neural fold homeobox-1* (*RX1*) is the candidate gene on LG23 [64]. The *RX1* gene is differentially expressed between the parents of the cross and is correlated in five cichlid species with sequenced genomes (R^2=0.64). A ~400bp deletion in the promoter located 2.5kb upstream of the start site was implicated as the causative mutation. Quantitative PCR analysis has shown that the presence of the deletion is correlated with decreased *RX1* expression, consequently decreasing *SWS2A* expression in most, if not all Lake Malawi cichlids. Specifically, the deletion explains 62% of the variation among 53 species tested.

The primary goal of this study is to evaluate candidate genes and identify the causative mutation contributing to differential *SWS2A* expression on LG5. This would complete the picture of *SWS2A* regulation in the retinas of African cichlids. Additionally, it would provide a peek into the under-studied cone photoreceptor specification network in a naturally occurring population.

5.3 MATERIALS AND METHODS

5.3.1 Ethics statement and sampling

All animals used in this work were bred and collected as part as several other previously published studies. In each case, animal procedures were approved by the Institutional Animal Care and Use Committee of University of Maryland, College Park (protocol R-09-73). All the cichlids used in the study were sexually mature adults (>6 months) and were sacrificed with a lethal dose of 250mg/L buffered MS-222. Retinas and fin clips were dissected from the euthanized animals and stored at -20^{o} in RNAlater and 100% ethanol respectively. To maintain consistency, all dissections were performed at the same time of the day.

5.3.2 Fine-mapping of LG5 QTL

Following previous methods, F2 hybrids (n=304) from the TA cross were genotyped using 27 microsatellite markers interspersed at regular intervals on LG5 (linkage groups are numbered according to the riverine ancestor *Oreochromis niloticus,* which has a high quality genome assembly) [64]. The microsatellite markers were designed using the *Metriaclima zebra* Broad assembly which is the most closely related genome to the parents

of the TA cross and is the reference genome for Lake Malawi cichlids [152]. DNA from the fin clips of F2 hybrids was extracted using Qiagen DNeasy Blood and Tissue Kit adopting manufacturer's protocol. After PCR amplification, genotyping reactions were run on an Applied Biosystems 3730xl and genotypes were inferred using ABI Genemapper software. Single marker regression analysis (association between marker genotype and opsin expression) were performed with a significance level $P < 0.05$ in R.

5.3.3 Retinal transcriptomes

Retinal transcriptomes from the five Broad genomes and the parents were analyzed to further evaluate candidate genes. This was based on the assumption that potential candidates would be differentially expressed between the parents of the cross and also correlated with *SWS2A* expression across the five cichlid species- *Oreochromis niloticus* (riverine ancestor), *Astatotilapia burtoni* (riverine), *Neolamprologus brichardi* (Lake Tanganyika), *Pundamilia nyererei* (Lake Victoria) and *Metriaclima zebra* (Lake Malawi). While three of the five reference species- *A. burtoni*, *P. nyererei* and *O. niloticus* have high *SWS2A* expression, *M. zebra* and *N. brichardi* have *SWS1* expression with no significant expression of *SWS2A* in their retinas. The construction and assembly of the retinal transcriptomes from the parental species was described earlier in Schulte et al, 2014 [64]. Succinctly, pooled retinal libraries from two *T. intermedius* and four *A. baenschi* were sequenced on an Illumina HiSeq 1000. Transciptome reads were aligned against the *M. zebra* genome and analyzed using Cufflinks and Cuffdiff to determine FPKM values for each of the NCBI Refseq gene models from *O. niloticus*. Retinal transcripts from the F1 hybrid of the TA cross were also constructed using these methods.

For genes within the 95% confidence interval of the QTL on LG5, correlation coefficients (R) with *SWS2A* across the five cichlid genomes were calculated. Similarly differential expression for each of the genes within this region defined as the ratio of expression (F) between the parents (*T.intermedius*/*A. baenschi* if T>A or *A.baenschi*/*T.intermedius* if A>T) were calculated to assess the best candidates.

5.3.4 Quantitative PCR of candidate genes

Dual-labelled Taqman primers and probes specific to *MITFA*, *BHE40*, *FOXP1B*, *GNAT2* and *SWS2A* were used to perform quantitative real-time PCR in the F2 hybrids of the TA cross (n =30). Expression of each gene was normalized to the expression of a housekeeping gene, *GNAT2* - the cone specific alpha subunit of transducin, which as part of the photoreceptor transduction cascade is consistently highly expressed in the cone cells of the retina [67]. RNA was extracted from dissected F2 retinas (n=30) using the Qiagen QIAshredder and RNeasy Mini Kit and transcribed into cDNA with Invitrogen Superscript III following manufacturer instructions. 0.5ul of cDNA from each sample was quantified on a Roche Light cycler 480 for 40 cycles [95ºC (10s)/ 55ºC (30s)/ 65ºC (60s)]. Expression relative to *GNAT2* was calculated (in percentage) as:

$$T_i/T_{GNAT2} = (1+E_{GNAT2})^{Ct GNAT2} /(1+E_i)^{Cti}$$

where E_i is the PCR efficiency and C_{ti} is the critical cycle number for each gene calculated from replicate RT-PCR reactions respectively. RT-PCR primers and probes are listed in Table 5.1.

5.3.5 Comparisons of coding sequences of candidate genes

The coding sequences for each of the candidate genes, *BHE40*, *MITFA* and *FOXP1B* were obtained from the *A. baenschi* and *T. intermedius* transcriptomes. After translation using the Expasy tool, the amino acid sequences of each gene were aligned using Clustalw before making comparisons [165, 259]. Attention was paid to sites located within the annotated functional domains of each transcription factor.

Table 5.1: Taqman primer and probe sequences for the candidates on LG5 and housekeeping gene *GNAT2*.

Gene	Forward Primer 5'-3'	Reverse Primer 5'-3'	Probe 5'-3'
MITFA	GAAGACGGAGATTCAACATCAATG	GGCTTTCAGAATAGTGCCCTTATT	TACCCAAGTCAAATGATCCAGACATGCG
BHE40	AAGAGACGCGACAGAATAAACGA	CCCAGCGTTGTAAGCTTAAGATG	TGCATTGCCCAACTGAAAGATTTATTGCC
FOXP1B	CGGCACCGCCAAAGAAC	TGGATCCTTCTCTCTTGTGACTCA	AACCAGCATGCCTCCACCAACGG
GNAT2	CCAAGAAGATGCTGATAAGGAGTCT	AACGATGGTGCTCTTCCCC	TCAAGCTATTACTGCTTGGTGCTGGTGAGTC

5.3.6 Transcription factor binding site and allele-specific expression

Promoters (5 kb upstream of the start site) flanking the *SWS2A* opsin and the neighboring *SWS2B* and *LWS* opsins on LG5 as well the *RX1* gene on LG23 were inspected for binding sites for *FOXP1B*, *MITFA* and *BHE40* transcription factors.

If *cis*-acting mutations of the candidate genes are responsible for their differences in expression, then differential expression of *A. baenschi* and *T. intermedius* alleles should persist in a common *trans* environment in the F1 hybrid [137]. To test this, we identified

polymorphic sites in the coding sequence of two candidate genes, *MITFA* and *BHE40* in the F1 hybrid transcripts. The coverage ratio of the SNPs corresponding to the *A. baenschi* and *T. intermedius* allele in the F1 hybrids was compared to the ratio of coverage at the same site in the F0 parents to get the relative contribution of *cis* and *trans* regulation for each candidate gene. This can be calculated from

$$\%cis = \frac{log_2 \frac{A_{F1}}{T_{F1}}}{log_2 \frac{A_{F0}}{T_{F0}}}$$

where A and T is the relative expression of the *A. baenschi* or *T. intermedius* alleles and the subscript indicates where the expression is measured in the F0 or F1 generation.

5.3.7 Sequencing parental genomes of TA cross

Whole genomes of two *T. intermedius* and two *A. baenschi* individuals were sequenced to identify candidate genes on LG5 that might contribute to the differences in *SWS2A* expression. High molecular weight DNA was extracted from fin clips using the Qiagen DNeasy Blood and Tissue Kit. DNA concentration was quantified by fluorescence spectrophotometry and normalized for library preparation. Sequencing libraries for each species were constructed from pooled DNA of both individuals using the Illumina TruSeq DNA Sample Prep Kit. During construction, the pooled DNA was sheared to ~550bp using a Covaris Shearer and each individual within a library was indexed separately. A full lane of 100bp paired-end sequencing was performed for each species on an Illumina HiSeq 1500.

After quality checks and filtering with FastQC, raw reads from both lanes were pooled together and trimmed using Trimmomatic (*version 0.33*) with the following parameters: *ILLUMINACLIP:TruSeq2PE.fa:2:30:10:SLIDINGWINDOW:4LEADING:10TRAILING:10 CROP:100 HEADCROP:0 AVGQUAL:20 MINLEN:80* [260]. The trimmed reads were aligned to the *M.zebra* reference genome using the default settings on BWA mem (*version 0.7.12*) [161, 261]. Duplicates were marked and sorted with Picard (*version 2.1.0*) and assembly accuracy metrics were calculated using MgsMetrics. Species alignments were viewed in IGV and edited manually in Geneious (*version 11.0.5*) before evaluating consensus sequences [166, 262].

5.3.8 Sequencing the *MITFA* gene

The promoter and introns of the *MITFA* gene was Sanger sequenced with overlapping primer sets in both the parents of the cross as well in a panel of 16 species from Lake Malawi with previously characterized opsin expression profiles [64]. The primers were designed based on *M. zebra* genome and are provided in Table 5.2 [152]. DNA from each sample was PCR amplified and sequenced using Applied Biosystems Big Dye Terminators following size selection by gel-electrophoresis on a ABI 3730xl. Species comparisons were made after assembling the sequences for all the primers and editing manually.

A previously generated Lake Malawi specific repeat sequence database was used to inspect the sequence content of the insertion in the *MITFA* gene. This database was constructed by combining consensus repeat sequences generated by RepeatModeler (version open- 1.0.8) on the *M. zebra* genome with the Repbase RepeatMasker library version 20140131 [161,

263]. The insertion was blasted against this database to identify any potential transposable elements.

5.3.9 Two locus interaction analysis

To understand how *RX1* and *MITFA* genes might work together to modulate the expression of *SWS2A* opsin, F2 hybrids (n=120) of the TA cross were genotyped for the deletion (~ 413bp) in the *RX1* and insertion (~1400bp) in the *MITFA* locus respectively. Following PCR amplification with primers flanking both mutations, each F2 hybrid was genotyped at both loci by size using gel electrophoresis. Subsequently, the interaction between the *RX1* gene (LG23) and the *MITF* gene (LG5) was tested by plotting the genotypes at both the loci with *SWS2A* expression.

Similarly, an additional 35 species from Lake Malawi (with previously quantified opsin expression) were genotyped for the presence or absence of the insertion in the promoter of the *MITFA* gene. These genotypes were combined with the *RX1* deletion genotypes for the same species from Schulte et al., 2014 and placed in the context of a species tree of Lake Malawi [64].

5.4 RESULTS AND DISCUSSION

5.4.1 Fine-mapping of LG5 QTL and identification of candidate genes

The QTL on LG5 is the smaller of the two QTLs regulating *SWS2A* expression in the TA cross. Since the QTL is located ~720kb (~1 cM) away from the *SWS2A-SWS2B-LWS* tandem array on LG5, we speculate that it most likely is acting in *trans* to the opsin. This

QTL was fine-mapped with 27 microsatellite markers spanning 28Mb in 302 F2 hybrids. The 95% confidence interval mapped to a 3.6 Mb region that also includes the *SWS2A-SWS2B-LWS* opsin array. Even though the opsin genes are in the 95% confidence interval, previous studies have concluded that the proximal promoters of the opsins are mostly conserved and thus could not explain the expression differences [96]. The most significant marker 41846 having a LOD score of 10.55 and explains 13.56% of the variation in expression of *SWS2A* in the F2 progeny.

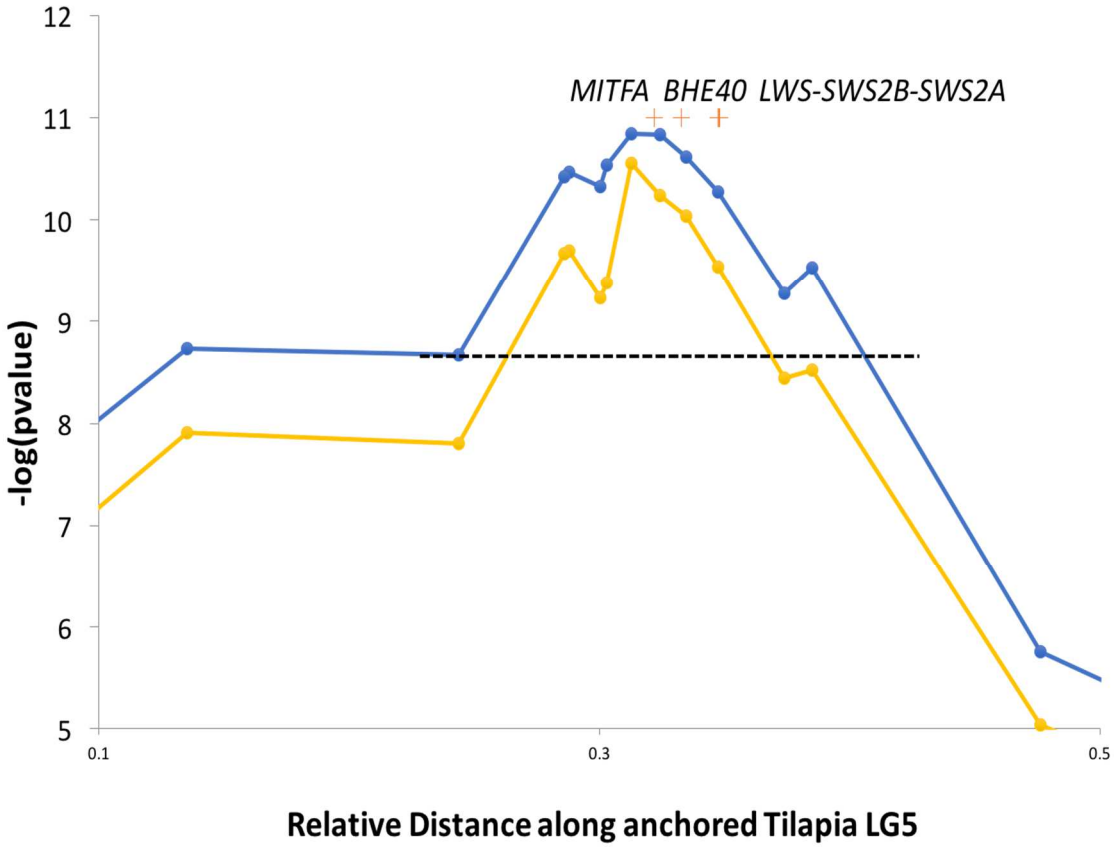


Fig. 5.1: Fine-mapping of the QTL on LG5. The QTL on LG5 was fine-mapped with microsatellite markers (indicated by the dots). The significance of association for each marker with SWS2A opsin (blue trace) and SWS2B opsin (yellow trace) is shown. The dotted horizontal line indicates the 95% confidence interval of the QTL. The relative distance in cM anchored to the Tilapia genome assembly is given on the x-axis. The candidate genes *BHE40*, *MITFA*, *FOXP1B* and the *SWS2A-SWS2B-LWS* opsin tandem array are all located in the 95% CI of the QTL. The + sign indicates the location of the genes.

Table 5.2: Sequencing primers for *MITFA* gene and promoter.

Primer Name	Forward primer 5'-3'	Reverse Primer 5'-3'
MITF-P1	AGACAGGTGTCTGTGAACGTG	CGGAAATATTTGTGGGTGCT
MITF-P2	AAAGGCTGGACTCTGCAGTTA	GCGAGTCAGTAGCAGAAGTACAGTAG
MITF-P3	CATCTCATGTGACGGCCACT	CCTTGTTGGTGCAAACGTATT
MITF-P4	AATACGTTTGCACCAACAAGG	GTGACCCCTTATTGGCCCTT
MITF-P5	TGCTGTTTCCCCATCTTTGT	GCTTGCTTAAGGCCAACTAACA
MITF-P6	TGTGTTTTGGGTGTTCAGAACA	ACGTGCTTGCTTTGCAACAT
MITF-P7	GTTGGACAGGTGCTGAAATGT	ACATCGTCGTTGTAGCTTGAC
MITF-P8	AGACAGAAGAAGGACAACCACA	TACGCAACATGAGATGCCGA
MITF-coding1	TGCTGGAGATGCTTGAATACAG	TCTCACAGTTGGTGCTGAGG
MITF-coding2	AGCAGTGCACCCAATAGTCC	TCGTCGTTGTAGCTTGACTCC

Based on the 95% confidence interval, candidate genes in the 3.6 Mb region on LG5 were explored using retinal transcriptomes of both the parents of the TA cross as well as the five cichlid genomes. The critical 3.6Mb interval contained a total of 106 genes including the *SWS2A-SWS2B-LWS* opsin array. Out of these 106 genes, the best candidates were filtered based on two criteria. First, the gene was differentially expressed by more than two-fold between the two grandparents- *A. baenschi* and *T. intermedius*. Second, expression of the gene was highly correlated (>30%) in the five Broad genomes [152]. This approach reduced the candidate list to 11 genes that were both differentially expressed in the cross and highly correlated across the Broad genomes. The list of these 11 genes including their functions are given in Table 5.3. Only three of the 11 genes are transcription factors with known roles in retinal regulation and development. These three genes include *Basic Helix-loop-Helix 40* (*BHE40*), *Microphthalmia-Associated Transcription Factor-A* (*MITFA*) and *Forkhead Box P1-B* (*FOXP1B*). Both *BHE40* and *MITFA* belong to a large family of basic helix-loop helix transcription factors that have conserved roles in retina embryogenesis in mice and zebrafish [264-266]. *FOXP1* transcription factors have been shown to mediate

cell cycle control in retinal precursor cells and also moderate cell fate determination during retinal development [267, 268]. Given the established role of these transcription factors in the retina, we focused our further efforts on these three genes.

Table 5.3: List of genes in the 95% CI on LG5 that were differentially expressed between the F0 of the cross and correlated with *SWS2A* expression.

Name of gene	Function
Phosphatase 1 regulatory subunit 3D	Participates in glycogen metabolism and protein synthesis.
Basic Helix-loop-Helix 40	Retina embryogenesis
Bactericidal permeability increasing protein1	Endogenous antibiotic protein
Forkhead Box P1-B	Cell cycle control in retinal precursor cells
Contactin-4	Cell-adhesion molecules for neural network formation.
Macrophage stimulating 1- Receptor	Involved in immune response and host defense.
Sulfatase Sulf-2	Co-receptors for growth factors and cytokines and involved in cell-signalling.
Ubiquitin CYLD	Regulation of microtubule dynamics
Cadherin 26	Cell-adhesion molecules associated with cell migration.
G-protein coupled receptor 22	Regulates cilia length in Kupffer's vesicle.
GATA-binding factor 2	Proliferation of hematopoietic and endocrine cell lineages.

For genes, all information was obtained from Genecards (http://www.genecards.org/).

5.4.2 Coding sequences of candidate genes

The three candidate genes identified in the study are transcription factors with well-conserved roles in other aspects of development and physiology. For example, while *BHE40* plays a role in the circadian cycle regulating the activity of the clock genes, *MITF* is well-known to be involved in melanocyte and retinal pigmented epithelium regulation [269, 270]. *FOXP1* transcription factors are ubiquitous in development, being involved in the differentiation of the lung epithelium and proliferation of cardiac cells and spinal motor neurons among many others [271, 272]. As such, any coding sequence mutations in these genes might be detrimental to the survival of the organism.

Comparisons of the *A. baenschi* and *T. intermedius* amino acid sequences for each of the three genes showed that a few sites do vary between the two species. Nevertheless, the major functional domains of the transcription factors were fully conserved between them. Thus, we suspect that the three candidate transcription factors are fully operative in the retinas of both the species.

5.4.3 Correlation between candidate genes and *SWS2A* expression

In addition to being associated with *SWS2A* expression in the transcriptomes, the "best" candidate gene mediating its expression would also be correlated in the hybrids of the TA cross. We used real-time PCR to test the expression of each of the three candidates (normalized to cone-specific housekeeping gene *GNAT2*) with *SWS2A* expression in 30 F2 progeny. Of the three genes, only *BHE40* showed a significant positive correlation with *SWS2A* expression (R^2= 0.26). Both *FOXP1B* (R^2= 0.013) and *MITFA* (R^2= 0.005) were not correlated (Fig. 5.1). Since the causative gene should be correlated with *SWS2A* in the F2, this points to *BHE40* as the best candidate.

5.4.4 Transcription factor binding analysis and Allele specific expression

We searched the promoters of the *SWS2A-SWS2B-LWS* (approximately 5kb upstream to the end of the array) opsin array in *M. zebra* for potential binding sites for the three candidate transcription factors. The promoter of the *LWS* opsin was also considered as a control.

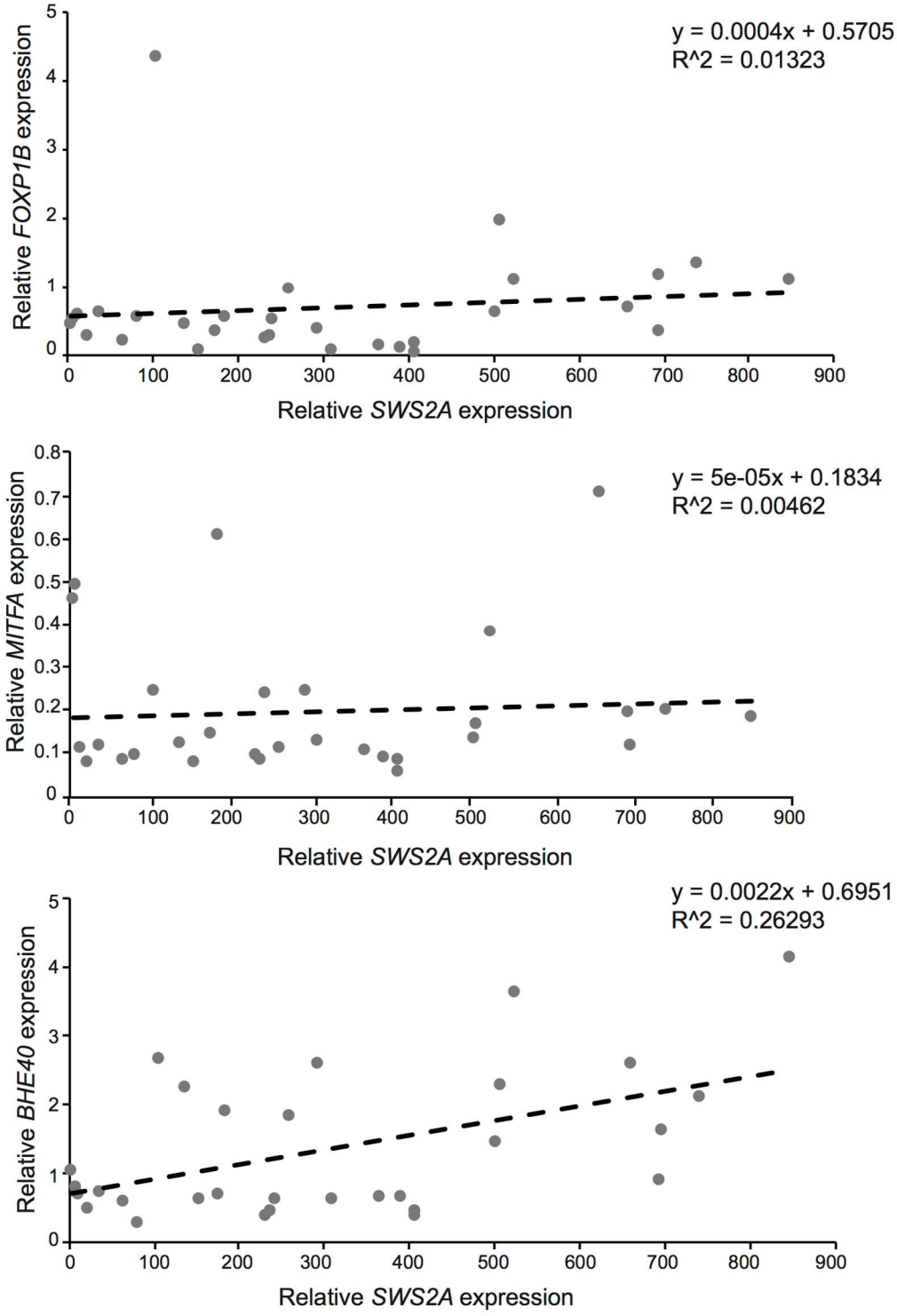


Fig. 5.2: Quantitative PCR of the three best candidate genes on LG5. Of the three candidates on LG5, only *BHE40* shows a positive correlation with *SWS2A* expression with F2 hybrids (n=30) from the cross. Expression of all candidates has been normalized to cone-specific housekeeping gene *GNAT2*.

Of the three candidates, *FOXP1B* (CGATACAA/AAGACAAC/AACACAAC) had the least number of putative binding sites with only one site in the promoter of the *SWS2B* opsin. *BHE40* (CACGTG/CACGCG/CACGAG) had three potential binding sites, with the *SWS2A* and *SWS2B* promoters having one and two binding sites respectively. The *SWS2A* promoter was enriched for *MITFA* (CACGTG/CATGTTG/TCATGTG) binding sites with three in the promoter. The *SWS2B* opsin, on the other hand, had two possible *MITFA* binding sites in the promoter.

Allele-specific expression calculations indicate that 100% of *MITFA* regulation exists in *cis*-configuration. However, the differential expression of *BHE40* between the two parents of the cross does not persist in the common *trans* environment in the F1 hybrid, with only 18.6% of the gene in *cis*-regulation. Since any variation in the expression of a causative transcription factor in the QTL must occur in cis to that gene, this suggests that *MITFA* could be the better candidate.

5.4.5 Comparisons of *cis*-regulatory sequences of the candidate genes

We sequenced the genomes of the grandparents, *T. interme*dius and *A. baenschi* of the cross to look for regulatory changes in the candidate genes. After Illumina sequencing, whole genomes from both species were mapped onto the *M. zebra* reference genome. While the two individuals of *T. intermedius* had a coverage of 13.04X $\pm$9.32 and 12.92X $\pm$ 9.07, the two *A. baenschi* sequenced had 10.96x $\pm$7.83 and 9.06X $\pm$6.68 coverage respectively. Around 5kb of the promoter upstream of the start site, all introns and 2kb of the 3' UTR was compared for each of the three genes.

The region of the genome encompassing *BHE40* had very low coverage in both species. A closer examination revealed that this region had low coverage even in the reference genome. Consequently, large gaps in sequence information were observed around this candidate, especially in the introns and 3' UTR of the gene in both the species, making species comparisons difficult. We were therefore unable to assess promoter variation. Further work will be required to use Sanger sequencing to target this region. For the *MITFA* gene, no large-scale differences between the two species were observed in either the promoter or the 3' UTRs. However, the 1st intron of the gene showed misassembly of sequences belonging to other linkage groups in *T. intermedius*. This misassembly was absent from both the reference *M. zebra* and *A. baenschi*.

5.4.6 An indel in the intron of *MITFA* is associated with *SWS2A* expression

In order to understand the discrepancy in the T. intermedius sequences in the 1st intron of the *MITFA* gene, the region was Sanger sequenced in both parents of the cross with overlapping primer sets. Sequencing analysis revealed a ~ 1400bp insertion in the first intron of the *MITFA* gene that was present exclusively in all the *T. intermedius* grandfathers and absent in *A. baenschi* grandmothers of the cross (Fig. 5.2). We then sequenced this region in a Lake Malawi cichlid panel comprising 16 species with previously quantified opsin expression profiles. A good association of allele size and opsin expression was observed among the panel. Similar to *Tramitochromis*, most species that have the insertion have high *SWS2A* opsin expression.

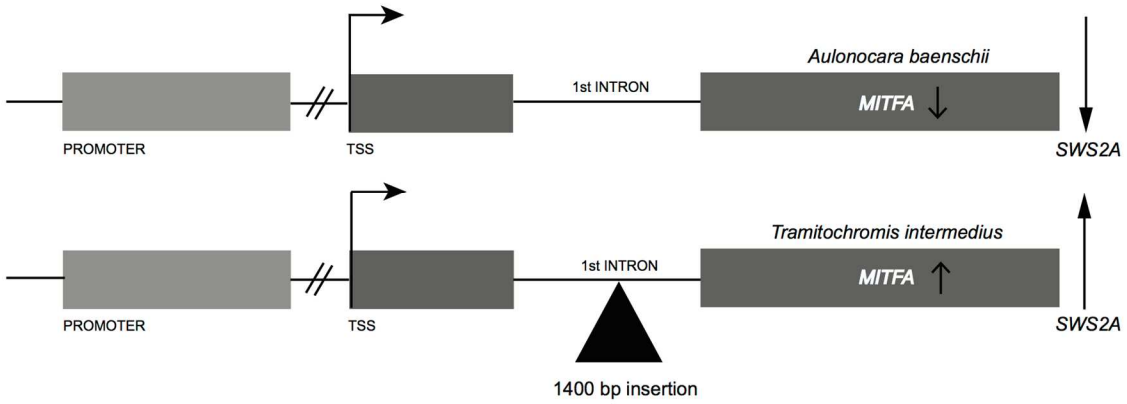


Fig. 5.3: Insertion in the first intron of the *MITFA* gene. Schematic representation of the *MITFA* gene in both the F0 of the TA cross. Sequencing revealed a 1400bp insertion located in the first intron of the *MITFA* gene in *T. intermedius*. This indel was correlated with increased *MITFA* expression causing an increase in *SWS2A* expression. The mutation is absent in *A. baenschi*.

The indel is restricted to Lake Malawi species and is absent from the riverine ancestor Tilapia or genomes of representative species from the other lakes. A closer inspection of the sequence content in the insertion using the cichlid-specific Repbase library revealed that it is highly repetitive and made up of two nested transposable elements, specifically a *REX-LINE* (*REX-1*) element overlapped on either side by a transposable element of unknown origin (*rnd-4_family 286*).

5.4.7 Interaction between *MITFA* and *RX1*

We then analyzed the *RX1* deletion and the *MITFA* insertion genotypes in 120 F2 individuals of the TA cross by size using gel electrophoresis. When based on size of fragment, the presence of the *RX1* deletion causes a smaller fragment (designated as the S allele) and the presence of the *MITFA* insertion results in a larger fragment (L allele) on the gel. This analysis revealed that the two genes work together to modulate the expression

of the *SWS2A* opsin, with the *RX1* locus exerting a stronger effect. Specifically, both genes have an additive effect on *SWS2A* expression. Absence of the deletion in *RX1* promoter (L allele) and the presence of the insertion in the first intron of the *MITFA* gene (L allele) seems to increase the expression of *SWS2A* in the F2 progeny (Fig. 5.3). The two-way-genotype table with the average expression for each genotype at both loci is provided in Table 5.4.

Table 5.4: Two-way genotype table with average expression of *SWS2A* for each genotype at both *MITFA* and *RX1* loci in the F2 hybrids.

	MITFA L/L	*MITFA* L/S	*MITFA* S/S
RX1 L/L	80.7	64.56	41
RX1 L/S	48.18	28.68	10.63
RX1 S/S	27.9	11.91	2.4

The same process was repeated for 53 Lake Malawi species whose opsin expression profiles are already known. In most species, presence of long (L) allele in both the genes seems to enhance the expression of *SWS2A*, with species homozygous for the L allele at both loci having the highest *SWS2A* opsin expression among species in the lake (Fig. 5.4). However, this analysis is incomplete because of the absence of species that are heterozygous for the *RX1* deletion. Future genotyping of additional species that are heterozygous for the *RX1* deletion but homozygous for the presence or absence of *MITFA* insertion will help make the relationship much clearer. None of the basal species in Lake Malawi and species outside the lake tested positive for the *MITFA* insertion. The two-locus interaction for the species panel is provided in Table 5.5.

Table 5.5: Two-way genotype table with average expression of *SWS2A* for each genotype at both *MITFA* and *RX1* loci in the Lake Malawi species panel. While *MITFA* is located on LG5, *RX1* is positioned on LG23. The number of species for each genotype group are given in brackets.

	MITFA L/L	*MITFA* L/S	*MITFA* S/S
RX1 L/L	68.5 (5)	83.01 (1)	15.48 (2)
RX1 L/S	(0)	56.89 (3)	(0)
RX1 S/S	5.11 (10)	4.85 (7)	4.88 (25)

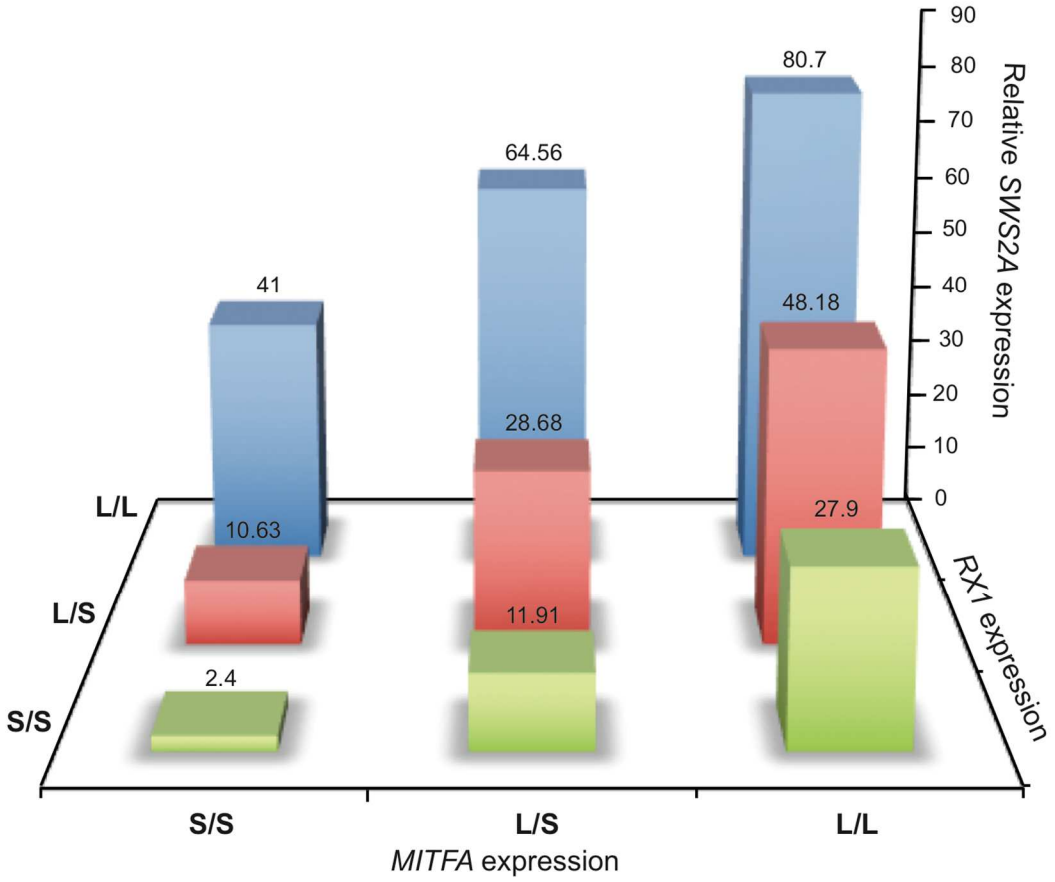


Fig. 5.4: Both *MITFA* and *RX1* work together to modulate the expression of *SWS2A*. The *RX1* locus on LG23 and *MITFA* gene on LG5 work additively to modulate the expression of *SWS2A* opsin in the F2 hybrids of the TA cross. F2 progeny homozygous for the absence of deletion at the *RX1* locus (L/L) and the presence of the insertion in the *MITFA* gene have the highest SWS2A expression. Based on size using gel electrophoresis- L/L: homozygous long fragment S/S: homozygous short fragment L/S: heterozygotes.

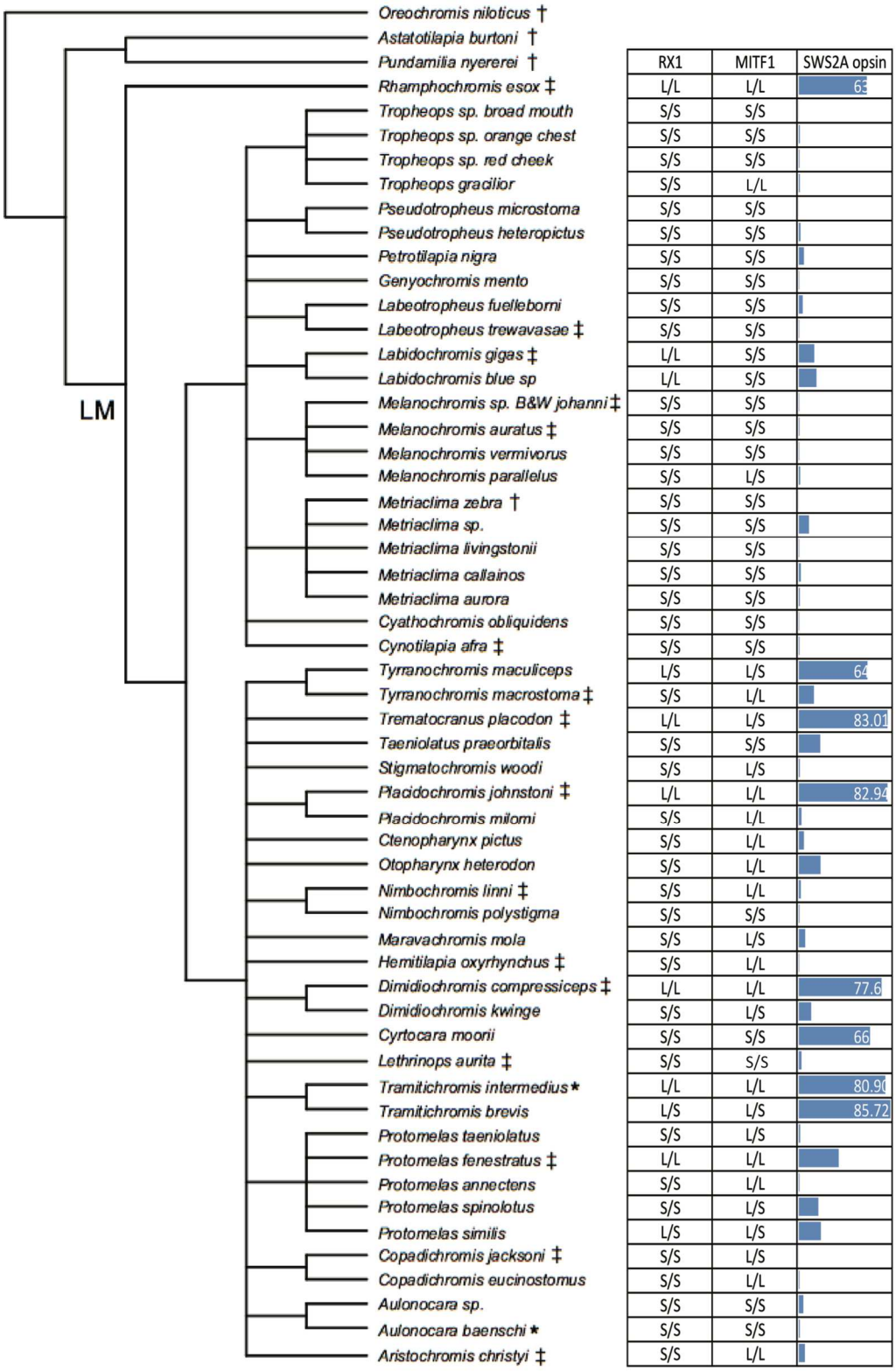

LM
Oreochromis niloticus †
Astatotilapia burtoni †
Pundamilia nyererei †
Rhamphochromis esox ‡
Tropheops sp. broad mouth
Tropheops sp. orange chest
Tropheops sp. red cheek
Tropheops gracilior
Pseudotropheus microstoma
Pseudotropheus heteropictus
Petrotilapia nigra
Genyochromis mento
Labeotropheus fuelleborni
Labeotropheus trewavasae ‡
Labidochromis gigas ‡
Labidochromis blue sp
Melanochromis sp. B&W johanni ‡
Melanochromis auratus ‡
Melanochromis vermivorus
Melanochromis parallelus
Metriaclima zebra †
Metriaclima sp.
Metriaclima livingstonii
Metriaclima callainos
Metriaclima aurora
Cyathochromis obliquidens
Cynotilapia afra ‡
Tyrranochromis maculiceps
Tyrranochromis macrostoma ‡
Trematocranus placodon ‡
Taeniolatus praeorbitalis
Stigmatochromis woodi
Placidochromis johnstoni ‡
Placidochromis milomi
Ctenopharynx pictus
Otopharynx heterodon
Nimbochromis linni ‡
Nimbochromis polystigma
Maravachromis mola
Hemitilapia oxyrhynchus ‡
Dimidiochromis compressiceps ‡
Dimidiochromis kwinge
Cyrtocara moorii
Lethrinops aurita ‡
Tramitichromis intermedius *
Tramitichromis brevis
Protomelas taeniolatus
Protomelas fenestratus ‡
Protomelas annectens
Protomelas spinolotus
Protomelas similis
Copadichromis jacksoni ‡
Copadichromis eucinostomus
Aulonocara sp.
Aulonocara baenschi *
Aristochromis christyi ‡
RX1
MITF1
SWS2A opsin
L/L L/L 63
S/S S/S
S/S S/S
S/S S/S
S/S L/L
S/S S/S
S/S S/S
S/S S/S
S/S S/S
S/S S/S
S/S S/S
S/S S/S
L/L S/S
L/L S/S
S/S S/S
S/S S/S
S/S S/S
S/S L/S
S/S S/S
S/S S/S
S/S S/S
S/S S/S
S/S S/S
S/S S/S
L/S L/S 64
S/S L/L
L/L L/S 83.01
S/S S/S
S/S L/S
L/L L/L 82.94
S/S L/L
S/S L/L
S/S L/L
S/S L/L
S/S S/S
S/S L/S
S/S L/L
L/L L/L 77.6
S/S L/S
S/S S/S 66
S/S S/S
L/L L/L 80.91
L/S L/S 85.72
S/S L/S
L/L L/L
S/S L/L
S/S L/S
L/S L/S
S/S L/S
S/S L/L
S/S S/S
S/S L/L

Fig. 5.5: Two-locus interaction for *SWS2A* expression in Lake Malawi species panel. The phylogenetic tree is based on generic relationships of cichlid species. To the right of the tree, the expression of the *SWS2A* opsin (depicted as blue bars) and the genotypes at the two loci- *RX1* and *MITFA* are shown for each species. For the *RX1* locus: L/L- homozygous for absence of deletion, S/S- homozygous for presence of deletion, L/S- heterozygous. For the *MITFA* locus: L/L- homozygous for the presence of insertion, S/S- homozygous for the absence of insertion, L/S- heterozygous.

5.5 DISCUSSION

In order to find the second gene regulating the differential expression of *SWS2A* opsin, we have fine-mapped the QTL on LG5 to localize the best candidate genes. It must be noted that the *SWS2A* tandem array is located within the 95% CI of this QTL and therefore the *cis*-control of the opsin gene cannot be completely ruled out. Detailed analysis of the *SWS2A* promoter in the future will help solve this issue. Fine-mapping results narrowed the QTL to a 3.6Mb region on LG5 that contained nearly 100 genes. Based on the differential expression between the two parent species of the cross- *Aulonocara* and *Tramitochromis*, and correlation with *SWS2A* expression in the five cichlid genomes, we have identified three genes- *BHE40*, *MITFA* and *FOXP1B* as the 'best' candidates. All three candidate genes are fully functional in both species of the cross and have conserved role in photoreceptor development and differentiation in other organisms.

Our hypothesis is that the causative gene on LG5 is moderated by *cis*-regulatory changes that result in a change in its expression. The gene in turn modulates *SWS2A* expression either by activating the expression of *SWS2A* or by repressing *SWS2B*. This modulation is achieved by the binding of the causative gene to specific sites located upstream of the promoter of either the *SWS2A* or *2B* gene. Based on results from quantitative PCR in the F2 hybrids, transcription factor binding sites and allele-specific expression analysis, we

rank these three candidates from 'least likely' to 'most likely' with regards to moderating divergent expression of *SWS2A* opsin. We present evidence for each of our case and argue that additional studies need to be carried out to truly determine the causative gene.

Least likely: *FOXP1B*

Expression of the gene is not correlated with *SWS2A* expression in the F2 hybrids and there is paucity of binding sites for this gene in the promoters of either *SWS2A* or *SWS2B* opsins. In testing for a correlation in the qPCR assay, we are testing for a linear correlation. However, it is not necessary that the expression between the candidate gene and *SWS2A* always be linear. For example, the transcription factor could act like a switch, where once its abundance crosses a certain threshold, could turn on the *SWS2A* opsin. We do not see much change in *FOXP1B* expression so neither a linear nor switch like variation is evident. In addition, we have measured the correlation in expression for each of these candidates in adult F2 retinas. It is plausible that this candidate works at different developmental stages and thus might be poorly correlated in fully differentiated adult photoreceptors.

Likely: *BHE40*

Of the three candidates, *BHE40* shows the only positive correlation with *SWS2A* expression in the F2 hybrids of the cross. Also, it is one of the most differentially expressed genes between the parent species in the 95% confidence interval. This transcription factor has multiple binding sites in the promoters of the opsins (1 in the *SWS2A* promoter and 2 in the *SWS2B* promoter). However, contrary to our expectations, results from allele-specific expression indicate that this gene is mostly controlled in *trans*. Therefore, it would be

unlikely to be the causative gene in the QTL where any *BHE40* expression variation should be explained by *cis*. Additionally, the genomic region around *BHE40* had extremely low coverage in both the reference as well as sequenced parent genomes making promoter and UTR sequence comparisons between the F0 species impossible. Future Sanger sequencing of this genomic region might help discern any differences between the two parents and test for or rule out potential causative mutations. Association mapping could then be used to test if any of those differences are functionally relevant.

Most Likely: *MITFA*

This transcription factor shows significant enrichment for binding sites in the *SWS2A* promoter (three sites) and is completed controlled in *cis*-regulation. Its expression is not linearly correlated with *SWS2A* expression in the F2 hybrids, but as mentioned previously could be correlated non-linearly or in a different developmental stage.

Perhaps the best evidence that likely implicates *MITFA* in moderating differential *SWS2A* expression is the presence of an insertion in the first intron of the gene that is present in species with high *SWS2A* expression. Additionally, two-loci interaction analysis with *RX1* loci genotypes on LG23 show that both the genes work additively to moderate *SWS2A* expression in the F2 hybrids of the cross. Matching the indel sequence to a cichlid-specific transposable element library indicated that a major part of the insertion belonged to a REX LINE transposon. LINE transposable elements are non-LTR retrotransposons that move around in the genome via the copy and paste mechanism [273]. Research on LINE transposable elements has proposed that by virtue of their movement, these elements can

accrue structural or regulatory functions in the host genomes. In fact, previous studies in cichlids using FISH techniques have shown that the REX LINE elements are found in euchromatic portions of the genome [274]. We postulate that the insertion in the first intron of the *MITFA* gene was the remnant of a past transposable element activity that now contributes to the regulation of the *MITFA* gene. The ability of constructs with transgenes containing different parts of the *MITFA* insertion to drive the expression of a reporter gene would be tested in the future to find functional sites within the insertion. These specific sites could then be knocked-out using CRISPR targets to conclusively implicate the role of *MITFA* gene in driving *SWS2A* expression.

With the exception of a few species, both the *MITFA* and *RX1* loci work additively to modulate *SWS2A* expression in Lake Malawi flock. For example, *Cyrtocara moori* has high *SWS2A* expression even though it has the deletion at the *RX1* locus and no insertion in *MITFA* (Fig. 5.4). Regardless, the causative mutations for *RX1* and *MITF1* occur in both the sand-dwelling and rock-dwelling clades of Lake Malawi suggesting that they arose early and are being distributed within the species in the lake. In cichlids, repeated evolution from sharing of ancestral alleles has been shown to be an important source of diversity [186, 275]. For instance, standing variation for coding sequence of the *LWS* opsin has been shown to be sorting within and among lakes [276]. Ongoing work in the lab is aimed at estimating the ages of both the *RX1* and *MITF1* mutations to determine if they are specific to the Lake Malawi flock.

5.6 CONCLUSION

Finding the causative gene among a pool of candidates is literally looking for a needle in a haystack. While we have concentrated our efforts on the genes in this QTL region, it is possible that the expression of *SWS2A* opsin is regulated by non-coding RNAs or microRNAs present within this interval. Although the proximal promoters of the opsins are well-conserved between the two species, it is well-known that *cis*-enhancers could be located several kb upstream of the gene they regulate. Future work will concentrate on comparing more upstream regions of the *SWS2A* opsin. While our efforts were focused on looking for large indel mutations, it is plausible that single nucleotide polymorphisms between the two species drive the differences in *SWS2A* expression. Additionally, with the exception of *BHE40*, most of the transcription factors in the LG5 region are very lowly expressed making quantitative PCR assays statistically noisy. The candidate gene is likely to be coexpressed in the same photoreceptor along with *SWS2A* expression. Future work to localize candidate gene to *SWS2A* photoreceptors in cichlid retinas might provide more clues. These could be done either using in-situ hybridization or single cell PCR.

Funding

This work was supported by the National Eye Institute at the National Institutes of Health (NIH R01EY024639 to K.L.C).

Author's Contributions

S.P.N performed all experiments pertaining to candidate gene evaluation and drafted the manuscript. J.E.S fine-mapped the QTL and identified candidate genes. B.A.S was responsible for grandparent genome assembly and M.A.C evaluated the repeat sequences bioinformatically. K.L.C conceptualized the study and critically reviewed the article.

Acknowledgements

We would like to thank Kelly O' Quin for all the groundwork on the TA cross and for leading the way, both literally and figuratively.

Chapter 6: Synthesis

6.1 INTRODUCTION

Deciphering the mechanisms governing phenotypic diversity is one of the central problems of evolutionary biology. Organismal phenotypes are influenced by underlying genetic changes but are also shaped by environmental inputs. While most studies focus on morphological divergence, sensory systems such as vision evolve through changes in genes and also by environmental modulation [212]. How are visual systems tuned by genetic factors? How does environmental plasticity cause visual system divergence?

The visual system in African cichlids is great system in which to study these questions. All African cichlids have seven distinct cone opsin genes for color vision [43]. Each species usually expresses a subset of these seven opsins, with species differing broadly in which opsins are expressed [49]. This differential expression of opsins leads to dramatic differences in the visual sensitivities among closely-related species, and has fitness consequences based on foraging, environmental-preference and mate choice [56, 59, 61, 62, 156]. Although prior work has indicated that the differential expression of opsins is regulated by genes as well as environmental-induced plasticity, the proximate mechanisms underlying both genetic and environmental control are not fully understood [50, 63, 64]. The work presented in this thesis has focused on trying to understand these mechanisms, the main results of which are summarized as follows.

6.2 MAIN INFERENCES

A few genes underlie differential cone opsin expression

Quantitative genetic analysis on the second-generation progeny of an intergeneric cross between two cichlid species with varying opsin expression- *Aulonocara baenschi* (*SWS1*, *RH2B*, *RH2A*) and *Metriaclima mbenji* (*SWS1*, *RH2B*, *RH2A*) (MA cross) has indicated that differential opsin expression in cichlids has a relatively simple genetic architecture, with only a handful of genes controlling expression differences [66]. The ease of opsin expression change between species and even between genera results in closely-related species with substantial variation in the spectral properties of photoreceptors, and could have facilitated species divergence. Additionally, the lability in opsin expression might have allowed cichlids to rapidly occupy new habitats, adapt to different foraging niches and survive photic environment changes- all hallmarks of a successful adaptive radiation.

Both *cis* and *trans* genetic changes underlie opsin expression variation

There have been animated debates in the literature about the relative importance of coding versus regulatory changes and *cis* versus *trans* changes to phenotypic divergence [72, 73, 134, 135, 187]. Many of these studies have suggested that *cis*-regulatory changes predominate owing to their modularity and lack of negative pleiotropy [133]. These studies relied more on morphological traits and we argue that detailed surveys of the genetic mechanisms underlying different kinds of organismal traits, such as those pertaining to physiology and behavior, need to be made before drawing any conclusions.

Results from a genome-wide mapping of the F2 progeny of the MA cross has revealed that opsin regulation in cichlids has evolved more through changes in *trans*-acting factors than *cis*-regulatory elements (Nandamuri et al., 2018, In Review). A total of 12 QTL controlling

the divergent expression of five cone opsins were identified, out of which only one for *SWS1* opsin was in cis-configuration (the remaining two QTL for *SWS1* were located in *trans* to the *SWS1* opsin). We speculate that for opsin expression divergence in cichlids, *trans* mutations might be favored over *cis* changes. Closely related cichlid species often exhibit coordinated differences in the expression of multiple opsins. Such synchronized changes can be more easily achieved by a single *trans* change influencing multiple downstream opsins as opposed to individual *cis* mutations in the opsin promoters.

Expression of *SWS1* opsin is controlled by a *cis*-regulatory indel

QTL analysis on the hybrids of the MA cross revealed a *cis*-QTL for differential expression of *SWS1* on LG17 (Nandamuri et al., 2018, In Review). This is the only example of a *cis*-regulatory QTL for opsin expression in African cichlids. A closer look at this major QTL revealed the presence of a ~700bp deletion in the promoter of the *SWS1* opsin (located 750bp from the start site) that is associated with a decrease in its expression. Of the two parental species used in the cross, *A. baenschi* has the deletion (low *SWS1* expression), while *M. mbenji* does not (high *SWS1* expression). The indel is also present in a few species in Lake Malawi with low *SWS1* expression levels, suggesting that it might be an ancestral polymorphism within the lake. Additionally, the indel harbors a microRNA *mir-729* and a non-coding element, both are which are conserved in over 300 million years of teleost evolution, portentous of their functional role in *SWS1* regulation. This is the second example of a regulatory deletion mediating differential opsin expression, suggesting that indels might be crucial to the evolution of visual diversity in African cichlids.

Adult cichlids with UV sensitivity in Lake Malawi are plastic

In response to changes in light environment, especially in the ultraviolet range, adult cichlids from Lake Malawi show significant plastic changes in *SWS1* opsin expression [67]. While *SWS1* expression decreased when wild-caught cichlids were housed under UV deficient lighting in the lab, *SWS1* levels went up when lab-reared cichlids were placed under broad-spectrum illumination designed to mimic lake light conditions. Moreover, these reciprocal expression changes were exceedingly fast, with few changes observed in as little as three days. Thus, rapid and reversible plasticity in opsin expression allows adult cichlids to respond to local changes in water clarity due to change of season, excessive rainfall or algal blooms.

6.3 FUTURE DIRECTIONS

Functional studies to assess the significance of the *cis*-regulatory deletion

In order to conclusively implicate the role of the *cis*-regulatory deletion in the promoter of *SWS1* gene in modulating its expression, future knockout studies will be performed. CRISPR targets designed against the *mi-729* and/or important transcription factor binding sites in the non-coding element are being made to test their role in regulating *SWS1* expression. Since CRIPSR technology has not yet been developed in Lake Malawi cichlids, these experiments will be performed in Tilapia. *O. niloticus* do not express *SWS1* opsin as adults but show *SWS1* expression during development (30dpf) [51]. Transgenic fish would be tested within the first 30 days after birth to test the effects of the loss of the *mi-729* or/and sites within the CNE on *SWS1* expression.

Gene regulatory networks for opsin expression

Future fine-mapping of the QTLs obtained in our study will lead to the identification of genes responsible for opsin expression divergence. Together with previously identified genes, the verified candidates for each opsin can be placed in the context of gene regulatory network of cichlid cone photoreceptor specification. This will provide exciting opportunities to draw hypotheses regarding the evolution of the network. Are certain genes more likely to accumulate mutations than others? Does the position of the gene within the network influence the kind of mutation it accumulates? How do mutations at one node of the network change the wiring within the network? Insights obtained by answering these questions can be expanded to infer the evolution of other complex traits across a broad range of phyla.

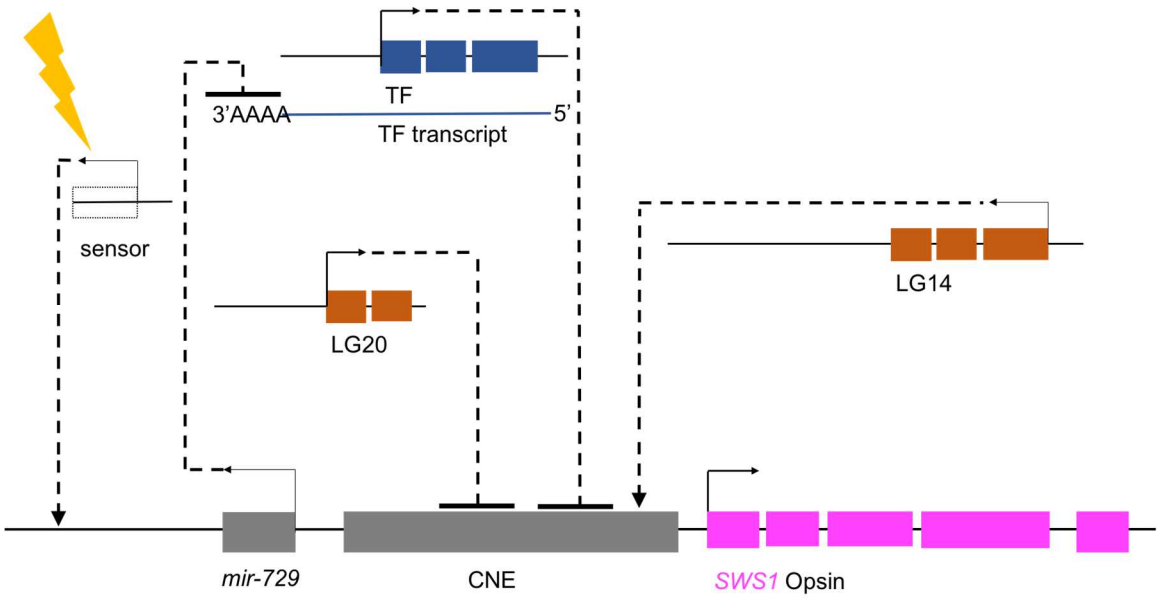


Fig. 6.1: Gene regulatory network for *SWS1* expression. *SWS1* expression in cichlids is genetically controlled by three QTL- LG14, LG20 and a *cis*-enhancer on LG17. The mi-RNA probably works as a negative-feedback loop to regulate *SWS1* expression. This might be important as the fish exhibits heterochronic shifts in opsin expression with age. Future fine-mapping will delineate the candidate genes on LG14 and LG20. *SWS1* expression is also influenced by UV light input. How exactly this change in light sensitivity is detected is not yet known. Horizontal lines indicate repression and arrows indicate activation.

QTL for plasticity

While *M. mbenji* show plastic changes in opsin expression between wild lake-caught and lab-reared specimens, *A. baenschi* never exhibit any light-induced plasticity, suggesting that plasticity to environmental light changes might have a genetic component [66]. Studies examining the genetic basis of plasticity, though traditionally debated, are now becoming more commonplace [233]. Splitting the F2 broods of the MA cross between UV-rich and UV-deficient environments might help pinpoint the genetic factors that contribute to plasticity in *SWS1* expression, given a common genetic environment. This will help us understand the evolution of plasticity in African cichlids a little better. Are the genes for plasticity the same as the ones that contribute to large genetic differences between the two species, or are they entirely different genes appropriated from the photoreceptor specification pathway?

REFERENCES

1. Darwin C: On the Origin of Species by Means of Natural Selection. Murray, London; 1858.
2. Pigliucci M, Murren CJ: Perspective: Genetic assimilation and a possible evolutionary paradox: Can macroevolution sometimes be so fast as to pass us by? *Evolution* 2003, 57(7):1455-1464.
3. Mayr E: Happy birthday: 80 years of watching the evolutionary scenery. *Science* 2004, 305(5680):46-47.
4. Waddington CH: Genetic Assimilation of an Acquired Character. *Evolution* 1953, 7(2):118-126.
5. Price TD, Qvarnstrom A, Irwin DE: The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society B-Biological Sciences* 2003, 270(1523):1433-1440.
6. Grant V: Organismic Evolution. San Franscisco: Freeman; 1977.
7. West-Eberhard MJ: Developmental Plasticity and Evolution. New York: Oxford University Press 2003.
8. West-Eberhard MJ: Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102:6543-6549.
9. Pigliucci M: Evolution of phenotypic plasticity: where are we going now? *Trends in Ecology & Evolution* 2005, 20(9):481-486.
10. Scheiner SM: Genetics and Evolution of Phenotypic Plasticity. *Annual Review of Ecology and Systematics* 1993, 24:35-68.
11. Bradshaw WE, Holzapfel CM: Genetic shift in photoperiodic response correlated with global warming. *Proc Natl Acad Sci U S A* 2001, 98(25):14509-14511.
12. Hoekstra HE: Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity (Edinb)* 2006, 97(3):222-234.
13. Sirkia PM, Virolainen M, Laaksonen T: Melanin coloration has temperature-dependent effects on breeding performance that may maintain phenotypic variation in a passerine bird. *J Evol Biol* 2010, 23(11):2385-2396.
14. Reznick D, Yang AP: The Influence of Fluctuating Resources on Life-History - Patterns of Allocation and Plasticity in Female Guppies. *Ecology* 1993, 74(7):2011-2019.
15. Reznick DN, Butler MJt, Rodd FH, Ross P: Life-History Evolution in Guppies (Poecilia Reticulata) 6. Differential Mortality as a Mechanism for Natural Selection. *Evolution* 1996, 50(4):1651-1660.
16. Wund MA, Baker JA, Clancy B, Golub JL, Fosterk SA: A test of the "Flexible stem" model of evolution: Ancestral plasticity, genetic accommodation, and morphological divergence in the threespine stickleback radiation. *American Naturalist* 2008, 172(4):449-462.
17. Bradbury JW, Vehrencamp SL: Principles of Animal Communication: Sinauer; 2011.
18. West-Eberhard MJ: Alternative adaptations, speciation, and phylogeny (A Review). *Proc Natl Acad Sci U S A* 1986, 83(5):1388-1392.

19. Wittkopp PJ: Variable gene expression in eukaryotes: a network perspective. *J Exp Biol* 2007, 210(Pt 9):1567-1575.
20. Wald G: Carotenoids and the visual cycle. *J Gen Physiol* 1935, 19(2):351-371.
21. Yokoyama S: Evolution of dim-light and color vision pigments. *Annual Review of Genomics and Human Genetics* 2008, 9:259-282.
22. Chang BS, Crandall KA, Carulli JP, Hartl DL: Opsin phylogeny and evolution: a model for blue shifts in wavelength regulation. *Mol Phylogenet Evol* 1995, 4(1):31-43.
23. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE *et al*: Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 2000, 289(5480):739-745.
24. Yokoyama S: Molecular evolution of vertebrate visual pigments. *Progress in Retinal and Eye Research* 2000, 19(4):385-419.
25. Yokoyama S, Yokoyama R: Adaptive evolution of photoreceptors and visual pigments in vertebrates. *Annual Review of Ecology and Systematics* 1996, 27:543-567.
26. Lythgoe J: Visual pigments and environmental light. *Vision research* 1984, 24(11):1539-1550.
27. Hunt DM, Bowmaker JK, Cowing JA, Carvalho LDS, Parry JWL, Wilkie SE, Davies WL: Spectral tuning of vertebrate visual pigments. *Perception* 2006, 35:167-167.
28. Hunt DM, Fitzgibbon J, Slobodyanyuk SJ, Bowmaker JK: Spectral tuning and molecular evolution of rod visual pigments in the species flock of cottoid fish in Lake Baikal. *Vision Res* 1996, 36(9):1217-1224.
29. Hunt DM, Carvalho LS, Cowing JA, Davies WL: Evolution and spectral tuning of visual pigments in birds and mammals. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2009, 364(1531):2941-2955.
30. Davies WL, Collin S, Hunt DM: Molecular ecology and adaptation of visual photopigments in craniates. *Molecular Ecology* 2012, 21(13):3121-3158.
31. Shand J, Davies WL, Thomas N, Balmer L, Cowing JA, Pointer M, Carvalho LS, Trezise AEO, Collin SP, Beazley LD *et al*: The influence of ontogeny and light environment on the expression of visual pigment opsins in the retina of the black bream, *Acanthopagrus butcheri*. *Journal of Experimental Biology* 2008, 211(9):1495-1503.
32. Fuller RC, Carleton KL, Fadool JM, Spady TC, Travis J: Genetic and environmental variation in the visual properties of bluefin killifish, *Lucania goodei*. *J Evol Biol* 2005, 18(3):516-523.
33. Fuller RC: Lighting environment predicts the relative abundance of male colour morphs in bluefin killifish (*Lucania goodei*) populations. *Proc R Soc Lond B Biol Sci* 2002, 269(1499):1457-1465.
34. Wagner HJ, Kroger RH: Adaptive plasticity during the development of colour vision. *Prog Retin Eye Res* 2005, 24(4):521-536.
35. Endler JA: Variation in the appearance of guppy color patterns to guppies and their predators under different visual conditions. *Vision Research* 1991, 31(3):587-608.
36. Osorio D, Vorobyev M: A review of the evolution of animal colour vision and visual communication signals. *Vision Research* 2008, 48(20):2042-2051.

37. Briscoe AD, Bybee SM, Bernard GD, Yuan F, Sison-Mangus MP, Reed RD, Warren AD, Llorente-Bousquets J, Chiao C: Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in *Heliconius* butterflies. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107(8):3628-3633.
38. Cummings ME: Sensory trade-offs predict signal divergence in surfperch. *Evolution* 2007, 61(3):530-545.
39. Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HD, Miyagi R, van der Sluijs I, Schneider MV, Maan ME, Tachida H *et al*: Speciation through sensory drive in cichlid fish. *Nature* 2008, 455(7213):620-U623.
40. Boughman JW: How sensory drive can promote speciation. *Trends in Ecology and Evolution* 2002, 17(12):571-577.
41. Kawata M, Shoji A, Kawamura S, Seehausen O: A genetically explicit model of speciation by sensory drive within a continuous population in aquatic environments. *BMC Evol Biol* 2007, 7:99.
42. Endler JA: Sensory drive does sensory biology bias or constrain the direction of evolution - Introduction to the symposium. *American Naturalist* 1992, 139:S1-S3.
43. Spady TC, Parry JWL, Robinson P, Hunt DM, Bowmaker JK, Carleton KL: Evolution of the cichlid visual palette through ontogenetic subfunctionalization of the opsin gene arrays. *Molecular Biology and Evolution* 2006, 23(8):1538-1547.
44. Bowmaker JK, Parry JW, Spady TC, Seehausen O, Hunt DM, Carleton KL: Divergent evolution and adaptation in cone visual pigments: Mix and match collour vision in African cichlid fish. *Perception* 2006, 35:116-117.
45. Parry JW, Carleton KL, Spady TC, Carboo A, Hunt DM, Bowmaker JK: Mix and match color vision: Tuning spectral sensitivity by differential opsin gene expression in Lake Malawi Cichlids. *Current Biology* 2005, 15(19):1734-1739.
46. Spady TC, Seehausen O, Loew ER, Jordan RC, Kocher TD, Carleton KL: Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Molecular Biology and Evolution* 2005, 22(6):1412-1422.
47. Smith AR, Carleton KL: Intrageneric Sequence Diversity in Cichlid Opsin Arrays. *Integrative and Comparative Biology* 2009, 49:E307-E307.
48. Smith AR, Carleton KL: Allelic Variation in Malawi Cichlid Opsins: A Tale of Two Genera. *Journal of Molecular Evolution* 2010, 70(6):593-604.
49. Carleton KL: Cichlid fish visual systems: mechanisms of spectral tuning. *Integrative Zoology* 2009, 4(1):75-86.
50. Carleton KL, Hofmann CM, Klisz C, Patel Z, Chircus LM, Simenauer LH, Soodoo N, Albertson RC, Ser JR: Genetic basis of differential opsin gene expression in cichlid fishes. *Journal of Evolutionary Biology* 2010, 23(4):840-853.
51. Carleton KL, Spady TC, Streelman JT, Kidd MR, McFarland WN, Loew ER: Visual sensitivities tuned by heterochronic shifts in opsin gene expression. *Bmc Biology* 2008, 6:22.
52. O'Quin KE, Smith AR, Sharma A, Carleton KL: New evidence for the role of heterochrony in the repeated evolution of cichlid opsin expression. *Evolution & Development* 2011, 13(2):193-203.

53. Hofmann CM, O'Quin KE, Smith AR, Carleton KL: Plasticity of opsin gene expression in cichlids from Lake Malawi. *Molecular Ecology* 2010, 19(10):2064-2074.
54. Smith AR, Ma K, Soares D, Carleton KL: Relative LWS cone opsin expression determines optomotor thresholds in Malawi cichlid fish. *Genes Brain and Behavior* 2012, 11(2):185-192.
55. Dalton BE, Lu J, Leips J, Cronin TW, Carleton KL: Variable light environments induce plastic spectral tuning by regional opsin coexpression in the African cichlid fish, *Metriaclima zebra*. *Molecular Ecology* 2015, 24(16):4193-4204.
56. Dobberfuhl AP, Ullmann JF, Shumway CA: Visual acuity, environmental complexity, and social organization in African cichlid fishes. *Behavioral Neuroscience* 2005, 119(6):1648-1655.
57. Smith AR, van Staaden MJ, Carleton KL: An evaluation of the role of sensory drive in the evolution of lake Malawi cichlid fishes. *International journal of evolutionary biology* 2012, 2012:647420.
58. O'Quin KE, Hofmann CM, Hofmann HA, Carleton KL: Parallel Evolution of Opsin Gene Expression in African Cichlid Fishes. *Molecular Biology and Evolution* 2010, 27(12):2839-2854.
59. Jordan R, Howe D, Juanes F, Stauffer J, Loew E: Ultraviolet radiation enhances zooplanktivory rate in ultraviolet sensitive cichlids. *African Journal of Ecology* 2004, 42(3):228-231.
60. Dijkstra PD, Seehausen O, Groothuis TGG: Direct male-male competition can facilitate invasion of new colour types in Lake Victoria cichlids. *Behavioral Ecology and Sociobiology* 2005, 58(2):136-143.
61. Kidd MR, Danley PD, Kocher TD: A direct assay of female choice in cichlids: all the eggs in one basket. *Journal of Fish Biology* 2006, 68(2):373-384.
62. Seehausen O, van Alphen JM: The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (*Haplochromis nyererei* complex). *Behavioral Ecology and Sociobiology* 1998, 42(1):1-8.
63. O'Quin KE, Schulte JE, Patel Z, Kahn N, Naseer Z, Wang H, Conte MA, Carleton KL: Evolution of cichlid vision via trans-regulatory divergence. *Bmc Evolutionary Biology* 2012, 12.
64. Schulte JE, O'Brien C, Conte MA, O'Quin KE, Carleton KL: Interspecific Variation in Rx1 Expression Controls Opsin Expression and Causes Visual System Diversity in African Cichlid Fishes. *Molecular Biology and Evolution* 2014, 31(9):2297-2308.
65. Hofmann CM, O'Quin KE, Smith AR, Carleton KL: Plasticity of opsin gene expression in cichlids from Lake Malawi. *Mol Ecol* 2010, 19(10):2064-2074.
66. Nandamuri SP, Dalton BE, Carleton KL: Determination of the genetic architecture underlying shortwavelength sensitivity in Lake Malawi cichlids. *Journal of Heredity* 2017, 108(4):379-390.
67. Nandamuri SP, Yourick MR, Carleton KL: Adult plasticity in African cichlids: Rapid changes in opsin expression in response to environmental light differences. *Mol Ecol* 2017, 26(21):6036-6052.

68. 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(7025):481-487.
69. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schluter D, Kingsley DM: Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 2004, 428(6984):717-723.
70. Clark RM, Wagler TN, Quijada P, Doebley J: A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. *Nat Genet* 2006, 38(5):594-597.
71. Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP: A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 2006, 313(5783):101-104.
72. Hoekstra HE, Coyne JA: The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 2007, 61(5):995-1016.
73. Carroll SB: Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 2008, 134(1):25-36.
74. Horth L: Sensory genes and mate choice: Evidence that duplications, mutations, and adaptive evolution alter variation in mating cue genes and their receptors. *Genomics* 2007, 90(2):159-175.
75. Collin SP, Davies WL, Hart NS, Hunt DM: The evolution of early vertebrate photoreceptors. *Philosophical Transactions of the Royal Society B-Biological Sciences* 2009, 364(1531):2925-2940.
76. Kocher TD: Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet* 2004, 5(4):288-298.
77. Seehausen O: African cichlid fish: a model system in adaptive radiation research. *Proc Biol Sci* 2006, 273(1597):1987-1998.
78. Salzburger W: The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology* 2009, 18(2):169-185.
79. Fryer G, & Iles TD: The cichlid fishes of the great lakes of Africa: Their biology and evolution.Oliver & Boyd, Croythorn House; 1972.
80. Genner MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, Turner GF: Age of cichlids: New dates for ancient lake fish radiations. *Molecular Biology and Evolution* 2007, 24(5):1269-1282.
81. Danley PD, Kocher TD: Speciation in rapidly diverging systems: lessons from Lake Malawi. *Molecular Ecology* 2001, 10(5):1075-1086.
82. Albertson RC, Markert JA, Danley PD, Kocher TD: Phylogeny of a rapidly evolving clade: The cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of Sciences of the United States of America* 1999, 96(9):5107-5110.
83. Albertson RC, Streelman JT, Kocher TD: Genetic basis of adaptive shape differences in the cichlid head. *Journal of Heredity* 2003, 94(4):291-301.
84. O'Quin CT, Drilea AC, Roberts RB, Kocher TD: A Small Number of Genes Underlie Male Pigmentation Traits in Lake Malawi Cichlid Fishes. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution* 2012, 318B(3):199-208.

85. Carleton KL, Kocher TD: Cone opsin genes of African cichlid fishes: Tuning spectral sensitivity by differential gene expression. *Molecular Biology and Evolution* 2001, 18(8):1540-1550.
86. Adler R, Raymond PA: Have we achieved a unified model of photoreceptor cell fate specification in vertebrates? *Brain Res* 2008, 1192:134-150.
87. Swaroop A, Kim D, Forrest D: Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. *Nature Reviews Neuroscience* 2010, 11(8):563-576.
88. Browman HI, Hawryshyn CW: Retinoic acid modulates retinal development in the juveniles op a teleost fish. *Journal of Experimental Biology* 1994, 193:191-207.
89. Picking WL, Chen D, Lee RD, Vogt ME, Polizzi JL, Marietta RG, Stark WS: Control of Drosophila opsin gene expression by carotenoids and retinoic acid: Northern and western analyses. *Experimental Eye Research* 1996, 63(5):493-500.
90. Li A, Zhu X, Brown B, Craft CM: Gene expression networks underlying retinoic acid-induced differentiation of human retinoblastoma cells. *Investigative Ophthalmology & Visual Science* 2003, 44(3):996-1007.
91. Roberts MR, Hendrickson A, McGuire CR, Reh TA: Retinoid X receptor gamma is necessary to establish the S-opsin gradient in cone photoreceptors of the developing mouse retina. *Investigative Ophthalmology & Visual Science* 2005, 46(8):2897-2904.
92. Chinen A, Hamaoka T, Yamada Y, Kawamura S: Gene duplication and spectral diversification of cone visual pigments of zebrafish. *Genetics* 2003, 163(2):663-675.
93. Tsujimura T, Chinen A, Kawamura S: Identification of a locus control region for quadruplicated green-sensitive opsin genes in zebrafish. *Proceedings of the National Academy of Sciences of the United States of America* 2007, 104(31):12813-12818.
94. Tsujimura T, Hosoya T, Kawamura S: A Single Enhancer Regulating the Differential Expression of Duplicated Red-Sensitive Opsin Genes in Zebrafish. *Plos Genetics* 2010, 6(12).
95. Mitchell DM, Stevens CB, Frey RA, Hunter SS, Ashino R, Kawamura S, Stenkamp DL: Retinoic Acid Signaling Regulates Differential Expression of the Tandemly-Duplicated Long Wavelength-Sensitive Cone Opsin Genes in Zebrafish. *Plos Genetics* 2015, 11(8).
96. O'Quin KE, Smith D, Naseer Z, Schulte JE, Engel SD, Loh YH, Streelman JT, Boore JL, Carleton KL: Divergence in cis-regulatory sequences surrounding the opsin gene arrays of African cichlid fishes. *Bmc Evolutionary Biology* 2011, 11(1):120.
97. Meyer A: Phylogenetic-relationships and evolutionary processes in East-African cichlid fishes. *Trends in Ecology & Evolution* 1993, 8(8):279-284.
98. Konings A: Malawi Cichlids in their Natural Habitat . 2007. *El Paso, TX, US: Cichlid Press*.
99. Moran P, Kornfield I, Reinthal PN: Molecular systematics and radiation of the haplochromine cichlids (Teleostei: Perciformes) of Lake Malawi. *Copeia* 1994:274-288.

100. Halstenberg S, Lindgren KM, Samagh SP, Nadal-Vicens M, Balt S, Fernald RD: Diurnal rhythm of cone opsin expression in the teleost fish *Haplochromis burtoni*. *Vis Neurosci* 2005, 22(2):135-141.
101. Lande R: The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* 1981, 99(3-4):541-553.
102. Lynch M & Walsh B: Genetics and analysis of quantitative traits. Sunderland, MA: Sinauer.; 1998.
103. Dalton BE, Loew ER, Cronin TW, Carleton KL: Spectral tuning by opsin coexpression in retinal regions that view different parts of the visual field. *Proceedings of the Royal Society B-Biological Sciences* 2014, 281(1797).
104. Barthel LK, Raymond PA: In situ hybridization studies of retinal neurons. *Vertebrate Phototransduction and the Visual Cycle, Pt B* 2000, 316:579-590.
105. Lee B, Lee W, Streelman JT, Carleton KL, Howe AE, Hulata G, Slettan A, Stern JE, Terai Y, Kocher TD: A second-generation genetic linkage map of tilapia (Oreochromis spp.). *Genetics* 2005, 170(1):237-244.
106. Fernald RD: Chromatic organization of a cichlid fish retina. *Vision Research* 1981, 21(12):1749-1753.
107. Beavis WD: QTL analyses:power, precision, and accuracy. In. Molecular Dissection of Complex traits: CRC Press, Boca Raton, FL.; 1998: 145-162.
108. Applebury ML, Antoch MP, Baxter LC, Chun LL, Falk JD, Farhangfar F, Kage K, Krzystolik MG, Lyass LA, Robbins JT: The murine cone photoreceptor: A single cone type expresses both S and M opsins with retinal spatial patterning. *Neuron* 2000, 27(3):513-523.
109. Parry JWL, Bowmaker JK: Visual pigment coexpression in guinea pig cones: A microspectrophotometric study. *Investigative Ophthalmology & Visual Science* 2002, 43(5):1662-1665.
110. Rajkumar P, Rollmann SM, Cook TA, Layne JE: Molecular evidence for color discrimination in the Atlantic sand fiddler crab, *Uca pugilator*. *Journal of Experimental Biology* 2010, 213(24):4240-4248.
111. Ogawa Y, Awata H, Wakakuwa M, Kinoshita M, Stavenga DG, Arikawa K: Coexpression of three middle wavelength-absorbing visual pigments in sexually dimorphic photoreceptors of the butterfly *Colias erate*. *Journal of Comparative Physiology a-Neuroethology Sensory Neural and Behavioral Physiology* 2012, 198(12):857-867.
112. Jacobs GH, Williams GA, Fenwick JA: Influence of cone pigment coexpression on spectral sensitivity and color vision in the mouse. *Vision Research* 2004, 44(14):1615-1622.
113. Otto SP, Jones CD: Detecting the undetected: Estimating the total number of loci underlying a quantitative trait. *Genetics* 2000, 156(4):2093-2107.
114. Westerbergh A, Doebley J: Morphological traits defining species differences in wild relatives of maize are controlled by multiple quantitative trait loci. *Evolution* 2002, 56(2):273-283.
115. Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM: The genetic architecture of divergence between threespine stickleback species. *Nature* 2001, 414(6866):901-905.

116. 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(4):987-1001.
117. Chen J, Tucker C, Woodford B, szel A, Lem J, Gianellaborradori A, Simon M, Bogenmann E: The human blue opsin promoter directs transgene expression in short-wave cones and bipolar cells in the mouse retina. *Proceedings of the National Academy of Sciences of the United States of America* 1994, 91(7):2611-2615.
118. Chiu M, Nathans J: A sequence upstream of the mouse blue visual pigment gene directs blue cone-specific transgene expression in mouse retinas. *Visual Neuroscience* 1994, 11(4):773-780.
119. Wang Y, Macke JP, Merbs SL, Zack DJ, Klaunberg B, Bennett J, Gearhart J, Nathans J: A locus-control region adjacent to the human red and green visual pigment genes. *Neuron* 1992, 9(3):429-440.
120. Hofmann CM, O'Quin KE, Marshall NJ, Cronin TW, Seehausen O, Carleton KL: The Eyes Have It: Regulatory and Structural Changes Both Underlie Cichlid Visual Pigment Diversity. *Plos Biology* 2009, 7(12).
121. Won YJ, Sivasundar A, Wang Y, Hey J: On the origin of Lake Malawi cichlid species: A population genetic analysis of divergence. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102:6581-6586.
122. Husemann M, Nguyen R, Ding B, Danley PD: A genetic demographic analysis of Lake Malawi rock-dwelling cichlids using spatio-temporal sampling. *Molecular Ecology* 2015, 24(11):2686-2701.
123. Hofmann CM, O'Quin KE, Marshall NJ, Carleton KL: The relationship between lens transmission and opsin gene expression in cichlids from Lake Malawi. *Vision Research* 2010, 50(3):357-363.
124. Albertson RC, Kocher TD: Genetic architecture sets limits on transgressive segregation in hybrid cichlid fishes. *Evolution* 2005, 59(3):686-690.
125. Albertson RC, Streelman JT, Kocher TD: Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences of the United States of America* 2003, 100(9):5252-5257.
126. Stelkens RB, Schmid C, Selz O, Seehausen O: Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. *BMC Evol Biol* 2009, 9:283.
127. Egger B, Sefc KM, Makasa L, Sturmbauer C, Salzburger W: Introgressive hybridization between color morphs in a population of cichlid fishes twelve years after human-induced secondary admixis. *J Hered* 2012, 103(4):515-522.
128. Rieseberg LH, Widmer A, Arntz AM, Burke JM: The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 2003, 358(1434):1141-1147.
129. Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB: The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell* 2008, 132(5):783-793.
130. Gruber JD, Vogel K, Kalay G, Wittkopp PJ: Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: frequency, effects, and dominance. *PLoS Genet* 2012, 8(2):e1002497.

131. Tirosh I, Reikhav S, Levy AA, Barkai N: A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* 2009, 324(5927):659-662.
132. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L: Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet* 2003, 35(1):57-64.
133. Wray GA: The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 2007, 8(3):206-216.
134. Stern DL, Orgogozo V: The loci of evolution: how predictable is genetic evolution? *Evolution* 2008, 62(9):2155-2177.
135. Wittkopp PJ, Haerum BK, Clark AG: Evolutionary changes in cis and trans gene regulation. *Nature* 2004, 430(6995):85-88.
136. Wang D, Sung HM, Wang TY, Huang CJ, Yang P, Chang T, Wang YC, Tseng DL, Wu JP, Lee TC *et al*: Expression evolution in yeast genes of single-input modules is mainly due to changes in trans-acting factors. *Genome Res* 2007, 17(8):1161-1169.
137. Wittkopp PJ, Haerum BK, Clark AG: Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat Genet* 2008, 40(3):346-350.
138. Domyan ET, Kronenberg Z, Infante CR, Vickrey AI, Stringham SA, Bruders R, Guernsey MW, Park S, Payne J, Beckstead RB *et al*: Molecular shifts in limb identity underlie development of feathered feet in two domestic avian species. *Elife* 2016, 5:e12115.
139. Shapiro MD, Bell MA, Kingsley DM: Parallel genetic origins of pelvic reduction in vertebrates. *Proc Natl Acad Sci U S A* 2006, 103(37):13753-13758.
140. McGaugh SE, Gross JB, Aken B, Blin M, Borowsky R, Chalopin D, Hinaux H, Jeffery WR, Keene A, Ma L *et al*: The cavefish genome reveals candidate genes for eye loss. *Nat Commun* 2014, 5:5307.
141. Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE: On the origin and spread of an adaptive allele in deer mice. *Science* 2009, 325(5944):1095-1098.
142. Carroll J, Rossi EA, Porter J, Neitz J, Roorda A, Williams DR, Neitz M: Deletion of the X-linked opsin gene array locus control region (LCR) results in disruption of the cone mosaic. *Vision Research* 2010, 50(19):1989-1999.
143. 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(1):94-98.
144. Alvarez-Delfin K, Morris AC, Snelson CD, Gamse JT, Gupta T, Marlow FL, Mullins MC, Burgess HA, Granato M, Fadool JM: Tbx2b is required for ultraviolet photoreceptor cell specification during zebrafish retinal development. *Proc Natl Acad Sci U S A* 2009, 106(6):2023-2028.
145. Raine JC, Hawryshyn CW: Changes in thyroid hormone reception precede SWS1 opsin downregulation in trout retina. *J Exp Biol* 2009, 212(17):2781-2788.
146. Kocher TD: Adaptive evolution and explosive speciation: The cichlid fish model. *Nature Reviews Genetics* 2004, 5(4):288-298.
147. Seehausen O: African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B-Biological Sciences* 2006, 273(1597):1987-1998.

148. Kornfield I, Smith PF: African cichlid fishes: Model systems for evolutionary biology. *Annual Review of Ecology and Systematics* 2000, 31:163-196.
149. Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL: How many species of cichlid fishes are there in African lakes? *Molecular Ecology* 2001, 10(3):793-806.
150. Joyce DA, Lunt DH, Genner MJ, Turner GF, Bills R, Seehausen O: Repeated colonization and hybridization in Lake Malawi cichlids. *Curr Biol* 2011, 21(3):R108-109.
151. Parnell NF, Streelman JT: Genetic interactions controlling sex and color establish the potential for sexual conflict in Lake Malawi cichlid fishes. *Heredity (Edinb)* 2013, 110(3):239-246.
152. Brawand D, Wagner CE, Li Y, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E *et al*: The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 2014, 513(7518):375.
153. Roberts RB, Ser JR, Kocher TD: Sexual Conflict Resolved by Invasion of a Novel Sex Determiner in Lake Malawi Cichlid Fishes. *Science* 2009, 326(5955):998-1001.
154. York RA, Patil C, Hulsey CD, Anoruo O, Streelman JT, Fernald RD: Evolution of bower building in Lake Malawi cichlid fish: phylogeny, morphology, and behavior. *Frontiers in Ecology and Evolution* 2015, 3(18).
155. Dijkstra PD, Seehausen O, Pierotti MER, Groothuis TGG: Male-male competition and speciation: aggression bias towards differently coloured rivals varies between stages of speciation in a Lake Victoria cichlid species complex. *Journal of Evolutionary Biology* 2007, 20(2):496-502.
156. Selz OM, Pierotti MER, Maan ME, Schmid C, Seehausen O: Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. *Behavioral Ecology* 2014, 25(3):612-626.
157. Maan ME, Hofker KD, van Alphen JM, Seehausen O: Sensory drive in cichlid speciation. *American Naturalist* 2006, 167(6):947-954.
158. Carleton KL: Quantification of Transcript Levels with Quantitative RT-PCR. *Methods Mol Biol* 2011, 772:279-295.
159. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA: Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 2008, 3(10):e3376.
160. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH: Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3-Genes Genomes Genetics* 2011, 1(3):171-182.
161. Conte MA, Kocher TD: An improved genome reference for the African cichlid, Metriaclima zebra. *BMC Genomics* 2015, 16:724.
162. Broman KW, Sen S: Guide to QTL Mapping with R/qtl. *Stat Biol Health* 2009:1-396.
163. Kosambi DD: The Estimation of Map Distances from Recombination Values. In: *DD Kosambi: Selected Works in Mathematics and Statistics.* Edited by Ramaswamy R. New Delhi: Springer India; 2016: 125-130.

164. Conte MA, Gammerdinger WJ, Bartie KL, Penman DJ, Kocher TD: A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics* 2017, 18(1):341.
165. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A: ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 2003, 31(13):3784-3788.
166. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al*: Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012, 28(12):1647-1649.
167. Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R: Whole Genome Sequences Of Malawi Cichlids Reveal Multiple Radiations Interconnected By Gene Flow. *bioRxiv* 2017.
168. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I: VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 2004, 32(Web Server issue):W273-279.
169. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Program NCS, Green ED, Sidow A, Batzoglou S: LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* 2003, 13(4):721-731.
170. Kumar S, Stecher G, Suleski M, Hedges SB: TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* 2017, 34(7):1812-1819.
171. Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H *et al*: JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2014, 42:142-147.
172. Kozomara A, Griffiths-Jones S: miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014, 42:68-73.
173. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS: MicroRNA targets in Drosophila. *Genome Biol* 2003, 5(1):R1.
174. Albertson RC, Powder KE, Hu Y, Coyle KP, Roberts RB, Parsons KJ: Genetic basis of continuous variation in the levels and modular inheritance of pigmentation in cichlid fishes. *Mol Ecol* 2014, 23(21):5135-5150.
175. Blake JA, Ziman MR: Pax genes: regulators of lineage specification and progenitor cell maintenance. *Development* 2014, 141(4):737-751.
176. Ochocinska MJ, Hitchcock PF: NeuroD regulates proliferation of photoreceptor progenitors in the retina of the zebrafish. *Mech Dev* 2009, 126(3-4):128-141.
177. Ratscho N, Scholten A, Koch KW: Expression profiles of three novel sensory guanylate cyclases and guanylate cyclase-activating proteins in the zebrafish retina. *Biochim Biophys Acta* 2009, 1793(6):1110-1114.
178. Wada Y, Sugiyama J, Okano T, Fukada Y: GRK1 and GRK7: unique cellular distribution and widely different activities of opsin phosphorylation in the zebrafish rods and cones. *J Neurochem* 2006, 98(3):824-837.
179. Rosenfeld PJ, Cowley GS, McGee TL, Sandberg MA, Berson EL, Dryja TP: A null mutation in the rhodopsin gene causes rod photoreceptor dysfunction and autosomal recessive retinitis pigmentosa. *Nat Genet* 1992, 1(3):209-213.

180. Mollema NJ, Yuan Y, Jelcick AS, Sachs AJ, von Alpen D, Schorderet D, Escher P, Haider NB: Nuclear receptor Rev-erb alpha (Nr1d1) functions in concert with Nr2e3 to regulate transcriptional networks in the retina. *PLoS One* 2011, 6(3):e17494.
181. Kloosterman WP, Steiner FA, Berezikov E, de Bruijn E, van de Belt J, Verheul M, Cuppen E, Plasterk RH: Cloning and expression of new microRNAs from zebrafish. *Nucleic Acids Res* 2006, 34(9):2558-2569.
182. Daido Y, Hamanishi S, Kusakabe TG: Transcriptional co-regulation of evolutionarily conserved microRNA/cone opsin gene pairs: implications for photoreceptor subtype specification. *Dev Biol* 2014, 392(1):117-129.
183. Chen K, Rajewsky N: The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 2007, 8(2):93-103.
184. Turner DL, Snyder EY, Cepko CL: Lineage-independent determination of cell type in the embryonic mouse retina. *Neuron* 1990, 4(6):833-845.
185. Cayouette M, Poggi L, Harris WA: Lineage in the vertebrate retina. *Trends Neurosci* 2006, 29(10):563-570.
186. Loh YH, Bezault E, Muenzel FM, Roberts RB, Swofford R, Barluenga M, Kidd CE, Howe AE, Di Palma F, Lindblad-Toh K *et al*: Origins of Shared Genetic Variation in African Cichlids. *Molecular Biology and Evolution* 2013, 30(4):906-917.
187. Wittkopp PJ: Genomic sources of regulatory variation in cis and in trans. *Cellular and Molecular Life Sciences CMLS* 2005, 62(16):1779-1783.
188. Travis J: Evaluating the adaptive role of morphological plasticity. *Ecological Morphology* 1994,99-122.
189. Pigliucci M: Phenotypic Plasticity: Beyond Nature and Nurture. Baltimore: Johns Hopkins University Press; 2001.
190. Stebbins GL: Variation and Evolution in Plants. New York: Columbia University Press; 1950.
191. Ehrenreich IM, Pfennig DW: Genetic assimilation: a review of its potential proximate causes and evolutionary consequences. *Annals of Botany* 2016, 117(5):769-779.
192. Pfennig DW, McGee M: Resource polyphenism increases species richness: a test of the hypothesis. *Philosophical Transactions of the Royal Society B-Biological Sciences* 2010, 365(1540):577-591.
193. Ghalambor CK, McKay JK, Carroll SP, Reznick DN: Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology* 2007, 21(3):394-407.
194. Yeh PJ, Price TD: Adaptive phenotypic plasticity and the successful colonization of a novel environment. *American Naturalist* 2004, 164(4):531-542.
195. Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek AP: Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution* 2010, 25(8):459-467.
196. Cahan SH, Julian GE, Rissing SW, Schwander T, Parker JD, Keller L: Loss of phenotypic plasticity generates genotype-caste association in harvester ants. *Current Biology* 2004, 14(24):2277-2282.

197. Shaw KA, Scotti ML, Foster SA: Ancestral plasticity and the evolutionary diversification of courtship behaviour in threespine sticklebacks. *Animal Behaviour* 2007, 73:415-422.
198. Gutteling EW, Riksen JAG, Bakker J, Kammenga JE: Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in *Caenorhabditis elegans*. *Heredity* 2007, 98(1):28-37.
199. Brewer MS, Carter RA, Croucher PJP, Gillespie RG: Shifting habitats, morphology, and selective pressures: Developmental polyphenism in an adaptive radiation of Hawaiian spiders. *Evolution* 2015, 69(1):162-178.
200. Losos JB, Creer DA, Glossip D, Goellner R, Hampton A, Roberts G, Haskell N, Taylor P, Ettling J: Evolutionary implications of phenotypic plasticity in the hindlimb of the lizard *Anolis sagrei*. *Evolution* 2000, 54(1):301-305.
201. Bowmaker JK: The visual pigments of fish. *Progress in Retinal and Eye Research* 1995, 15(1):1-31.
202. Loew ER, Lythgoe JN: The ecology of cone pigments in teleost fishes. *Vision Research* 1978, 18(8):715-722.
203. Levine JS, Macnichol EF: Visual pigments in teleost fishes - effects of habitat, microhabitat, and behavior on visual-system evolution. *Sensory Processes* 1979, 3(2):95-131.
204. Hunt DM, Carvalho LS, Cowing JA, Parry JWL, Wilkie SE, Davies WL, Bowmaker JK: Spectral tuning of shortwave-sensitive visual pigments in vertebrates. *Photochemistry and Photobiology* 2007, 83(2):303-310.
205. Lamb TD, Collin SP, Pugh EN, Jr.: Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup. *Nat Rev Neurosci* 2007, 8(12):960-976.
206. Bloch NI, Morrow JM, Chang BS, Price TD: SWS2 visual pigment evolution as a test of historically contingent patterns of plumage color evolution in warblers. *Evolution* 2015, 69(2):341-356.
207. Beaudet L, Hawryshyn CW. In: *Adaptive Mechanisms in the ecology of vision.* Edited by Archer S, Djamgoz MB, Loew E, Patridge JC, Vallerga S. Cornwall: Kluwer Academic Publishers; 1999: 413-437.
208. Collin SP, Shand J: Retinal sampling and the visual field in fishes. In: *Sensory Processing in the Aquatic Environment.* Edited by Collin S, Marshall N. New York: Springer Verlag; 2003: 139-169.
209. Flamarique IN: Temporal shifts in visual pigment absorbance in the retina of Pacific salmon. *Journal of Comparative Physiology a-Neuroethology Sensory Neural and Behavioral Physiology* 2005, 191(1):37-49.
210. Sandkam BA, Deere-Machemer KA, Johnson AM, Grether GF, Rodd FH, Fuller RC: Exploring visual plasticity: dietary carotenoids can change color vision in guppies (*Poecilia reticulata*). *Journal of Comparative Physiology a-Neuroethology Sensory Neural and Behavioral Physiology* 2016, 202(7):527-534.
211. Toomey MB, McGraw KJ: The Effects of Dietary Carotenoid Supplementation and Retinal Carotenoid Accumulation on Vision-Mediated Foraging in the House Finch. *Plos One* 2011, 6(6).
212. Fuller RC, Carleton KL, Fadool JM, Spady TC, Travis J: Genetic and environmental variation in the visual properties of bluefin killifish, *Lucania goodei*. *Journal of Evolutionary Biology* 2005, 18(3):516-523.

213. Temple SE, Veldhoen KM, Phelan JT, Veldhoen NJ, Hawryshyn CW: Ontogenetic changes in photoreceptor opsin gene expression in coho salmon (*Oncorhynchus kisutch*, Walbaum). *Journal of Experimental Biology* 2008, 211(24):3879-3888.
214. Shand J: Changes in the Spectral Absorption of Cone Visual Pigments during the Settlement of the Goatfish *Upeneus tragula* - the Loss of Red Sensitivity as a Benthic Existence Begins. *J Comp Physiol A* 1993, 173(1):115-121.
215. Cottrill PB, Davies WL, Semo M, Bowmaker JK, Hunt DM, Jeffery G: Developmental dynamics of cone photoreceptors in the eel. *Bmc Dev Biol* 2009, 9.
216. Cheng CL, Flamarique IN: Chromatic organization of cone photoreceptors in the retina of rainbow trout: single cones irreversibly switch from UV (SWS1) to blue (SWS2) light sensitive opsin during natural development. *Journal of Experimental Biology* 2007, 210(23):4123-4135.
217. Hoke KL, Evans BI, Fernald RD: Remodeling of the cone photoreceptor mosaic during metamorphosis of flounder (*Pseudopleuronectes americanus*). *Brain Behav Evolut* 2006, 68(4):241-254.
218. Shand J, Hart NS, Thomas N, Partridge JC: Developmental changes in the cone visual pigments of black bream *Acanthopagrus butcheri*. *Journal of Experimental Biology* 2002, 205(23):3661-3667.
219. Ehlman SM, Sandkam BA, Breden F, Sih A: Developmental plasticity in vision and behavior may help guppies overcome increased turbidity. *Journal of Comparative Physiology a-Neuroethology Sensory Neural and Behavioral Physiology* 2015, 201(12):1125-1135.
220. Tobler M, Coleman SW, Perkins BD, Rosenthal GG: Reduced opsin gene expression in a cave-dwelling fish. *Biol Letters* 2010, 6(1):98-101.
221. Wagner HJ, Kroger RHH: Effects of long-term spectral deprivation on the morphological organization of the outer retina of the blue acara (*Aequidens pulcher*). *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 2000, 355(1401):1249-1252.
222. Flamarique IN, Cheng CL, Bergstrom C, Reimchen TE: Pronounced heritable variation and limited phenotypic plasticity in visual pigments and opsin expression of threespine stickleback photoreceptors. *J Exp Biol* 2013, 216(Pt 4):656-667.
223. Fuller RC, Carleton KL, Fadool JM, Spady TC, Travis J: Population variation in opsin expression in the bluefin killifish, *Lucania goodei*: a real-time PCR study. *Journal of Comparative Physiology a-Neuroethology Sensory Neural and Behavioral Physiology* 2004, 190(2):147-154.
224. Fuller RC, Noa LA, Strellner RS: Teasing Apart the Many Effects of Lighting Environment on Opsin Expression and Foraging Preference in Bluefin Killifish. *American Naturalist* 2010, 176(1):1-13.
225. Fuller RC, Claricoates KM: Rapid light-induced shifts in opsin expression: finding new opsins, discerning mechanisms of change, and implications for visual sensitivity. *Molecular Ecology* 2011, 20(16):3321-3335.
226. Dalton BE, Cronin TW, Marshall JN, Carleton KL: The fish eye view: are cichlids conspicuous? *Journal of Experimental Biology* 2010, 213(13):2243-2255.
227. Ribbink AJ, Marsh BA, Marsh AC, Ribbink AC, Sharp BJ: A Preliminary Survey of the Cichlid Fishes of Rocky Habitats in Lake Malawi. *S Afr J Zool* 1983, 18(3):149-310.

228. Kelber A, Vorobyev M, Osorio D: Animal colour vision - behavioural tests and physiological concepts. *Biol Rev* 2003, 78(1):81-118.
229. Govardovskii VI, Fyhrquist N, Reuter T, Kuzmin DG, Donner K: In search of the visual pigment template. *Visual Neuroscience* 2000, 17(4):509-528.
230. Escobar-Camacho D, Marshall JN, Carleton KL: Behavioral color vision in a cichlid fish: Metriaclima benetos. *J Exp Biol* 2017, 220(Pt 16):2887-2899.
231. Wu DM, Schneiderman T, Burgett J, Gokhale P, Barthel L, Raymond PA: Cones regenerate from retinal stem cells sequestered in the inner nuclear layer of adult goldfish retina. *Investigative Ophthalmology & Visual Science* 2001, 42(9):2115-2124.
232. Wen XH, Shen LX, Brush RS, Michaud N, Al-Ubaidi MR, Gurevich VV, Hamm HE, Lem J, DiBenedetto E, Anderson RE *et al*: Overexpression of Rhodopsin Alters the Structure and Photoresponse of Rod Photoreceptors. *Biophys J* 2009, 96(3):939-950.
233. Parsons KJ, Concannon M, Navon D, Wang J, Ea I, Groveas K, Campbell C, Albertson RC: Foraging environment determines the genetic architecture and evolutionary potential of trophic morphology in cichlid fishes. *Molecular Ecology* 2016, 25(24):6012-6023.
234. Bergland AO, Genissel A, Nuzhdin SV, Tatar M: Quantitative trait loci affecting phenotypic plasticity and the allometric relationship of ovariole number and thorax length in *Drosophila melanogaster*. *Genetics* 2008, 180(1):567-582.
235. Smith AR, D'Annunzio L, Smith AE, Sharma A, Hofmann CM, Marshall NJ, Carleton KL: Intraspecific cone opsin expression variation in the cichlids of Lake Malawi. *Molecular Ecology* 2011, 20(2):299-310.
236. Vonlanthen P, Bittner D, Hudson AG, Young KA, Muller R, Lundsgaard-Hansen B, Roy D, Di Piazza S, Largiader CR, Seehausen O: Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* 2012, 482(7385):357.
237. Guildford SJ, Bootsma HA, Taylor WD, Hecky RE: High variability of phytoplankton photosynthesis in response to environmental forcing in oligotrophic Lake Malawi/Nyasa. *J Great Lakes Res* 2007, 33(1):170-185.
238. Seehausen O, vanAlphen J, Witte F: Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* 1997, 277(5333):1808-1811.
239. Muschick M, Barluenga M, Salzburger W, Meyer A: Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. *Bmc Evolutionary Biology* 2011, 11.
240. Wimberger PH: Plasticity of Fish Body Shape - the Effects of Diet, Development, Family and Age in 2 Species of *Geophagus* (Pisces, Cichlidae). *Biological Journal of the Linnean Society* 1992, 45(3):197-218.
241. Crispo E, Chapman LJ: Geographic variation in phenotypic plasticity in response to dissolved oxygen in an African cichlid fish. *Journal of Evolutionary Biology* 2010, 23(10):2091-2103.
242. Maruska KP, Fernald RD: Behavioral and physiological plasticity: Rapid changes during social ascent in an African cichlid fish. *Horm Behav* 2010, 58(2):230-240.
243. Pigliucci M, Murren CJ, Schlichting CD: Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology* 2006, 209(12):2362-2367.

244. Lande R: Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology* 2009, 22(7):1435-1446.
245. Hornsby MAW, Sabbah S, Robertson RM, Hawryshyn CW: Modulation of environmental light alters reception and production of visual signals in Nile tilapia. *Journal of Experimental Biology* 2013, 216(16):3110-3122.
246. Relyea RA: Costs of phenotypic plasticity. *American Naturalist* 2002, 159(3):272-282.
247. Callahan HS, Maughan H, Steiner UK: Phenotypic Plasticity, Costs of Phenotypes, and Costs of Plasticity Toward an Integrative View. *Ann Ny Acad Sci* 2008, 1133:44-66.
248. Lyons RP, Scholz CA, Cohen AS, King JW, Brown ET, Ivory SJ, Johnson TC, Deino AL, Reinthal PN, McGlue MM *et al*: Continuous 1.3-million-year record of East African hydroclimate, and implications for patterns of evolution and biodiversity. *Proceedings of the National Academy of Sciences of the United States of America* 2015, 112(51):15568-15573.
249. Ivory SJ, Blome MW, King JW, McGlue MM, Cole JE, Cohen AS: Environmental change explains cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. *Proceedings of the National Academy of Sciences of the United States of America* 2016, 113(42):11895-11900.
250. Olson-Manning CF, Wagner MR, Mitchell-Olds T: Adaptive evolution: evaluating empirical support for theoretical predictions. *Nat Rev Genet* 2012, 13(12):867-877.
251. Hauser FE, Chang BS: Insights into visual pigment adaptation and diversity from model ecological and evolutionary systems. *Curr Opin Genet Dev* 2017, 47:110-120.
252. Kumar JP: My what big eyes you have: how the *Drosophila* retina grows. *Dev Neurobiol* 2011, 71(12):1133-1152.
253. Cepko CL, Austin CP, Yang X, Alexiades M, Ezzeddine D: Cell fate determination in the vertebrate retina. *Proc Natl Acad Sci U S A* 1996, 93(2):589-595.
254. Gregory-Evans CY, Wallace VA, Gregory-Evans K: Gene networks: dissecting pathways in retinal development and disease. *Prog Retin Eye Res* 2013, 33:40-66.
255. Hao H, Kim DS, Klocke B, Johnson KR, Cui K, Gotoh N, Zang C, Gregorski J, Gieser L, Peng W *et al*: Transcriptional regulation of rod photoreceptor homeostasis revealed by in vivo NRL targetome analysis. *PLoS Genet* 2012, 8(4):e1002649.
256. Hu J, Wan J, Hackler L, Jr., Zack DJ, Qian J: Computational analysis of tissue-specific gene networks: application to murine retinal functional studies. *Bioinformatics* 2010, 26(18):2289-2297.
257. Carleton KL, Parry JW, Bowmaker JK, Hunt DM, Seehausen O: Colour vision and speciation in Lake Victoria cichlids of the genus *Pundamilia*. *Molecular Ecology* 2005, 14(14):4341-4353.
258. 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(14):4341-4353.

259. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al*: Clustal W and Clustal X version 2.0. *Bioinformatics* 2007, 23(21):2947-2948.
260. Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014, 30(15):2114-2120.
261. Li H: Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:13033997* 2013.
262. Thorvaldsdottir H, Robinson JT, Mesirov JP: Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* 2013, 14(2):178-192.
263. Smit AFA, Hubley R, Green P: RepeatModeler Open-1.0. 2008-2010. *Access date Dec* 2014.
264. Yao J, Wang L, Chen L, Zhang S, Zhao Q, Jia W, Xue J: Cloning and developmental expression of the DEC1 ortholog gene in zebrafish. *Gene Expr Patterns* 2006, 6(8):919-927.
265. Hallsson JH, Haflidadottir BS, Stivers C, Odenwald W, Arnheiter H, Pignoni F, Steingrimsson E: The basic helix-loop-helix leucine zipper transcription factor Mitf is conserved in *Drosophila* and functions in eye development. *Genetics* 2004, 167(1):233-241.
266. Nguyen M, Arnheiter H: Signaling and transcriptional regulation in early mammalian eye development: a link between FGF and MITF. *Development* 2000, 127(16):3581-3591.
267. Cheng L, Chong M, Fan W, Guo X, Zhang W, Yang X, Liu F, Gui Y, Lu D: Molecular cloning, characterization, and developmental expression of foxp1 in zebrafish. *Dev Genes Evol* 2007, 217(10):699-707.
268. Moose HE, Kelly LE, Nekkalapudi S, El-Hodiri HM: Ocular forkhead transcription factors: seeing eye to eye. *Int J Dev Biol* 2009, 53(1):29-36.
269. Bharti K, Gasper M, Ou J, Brucato M, Clore-Gronenborn K, Pickel J, Arnheiter H: A regulatory loop involving PAX6, MITF, and WNT signaling controls retinal pigment epithelium development. *PLoS Genet* 2012, 8(7):e1002757.
270. Kumasaka M, Sato S, Yajima I, Goding CR, Yamamoto H: Regulation of melanoblast and retinal pigment epithelium development by *Xenopus laevis* Mitf. *Dev Dyn* 2005, 234(3):523-534.
271. Zhang Y, Li S, Yuan L, Tian Y, Weidenfeld J, Yang J, Liu F, Chokas AL, Morrisey EE: Foxp1 coordinates cardiomyocyte proliferation through both cell-autonomous and nonautonomous mechanisms. *Genes Dev* 2010, 24(16):1746-1757.
272. Rousso DL, Gaber ZB, Wellik D, Morrisey EE, Novitch BG: Coordinated actions of the forkhead protein Foxp1 and Hox proteins in the columnar organization of spinal motor neurons. *Neuron* 2008, 59(2):226-240.
273. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O *et al*: A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 2007, 8(12):973-982.
274. Schneider CH, Gross MC, Terencio ML, do Carmo EJ, Martins C, Feldberg E: Evolutionary dynamics of retrotransposable elements Rex1, Rex3 and Rex6 in neotropical cichlid genomes. *BMC Evol Biol* 2013, 13:152.

275. Loh YH, Katz LS, Mims MC, Kocher TD, Yi SV, Streelman JT: Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. *Genome Biol* 2008, 9(7):R113.
276. Terai Y, Mayer WE, Klein J, Tichy H, Okada N: The effect of selection on a long wavelength-sensitive (LWS) opsin gene of Lake Victoria cichlid fishes. *Proc Natl Acad Sci U S A* 2002, 99(24):15501-15506.