ABSTRACT

Title of Dissertation: EVOLUTION, DEVELOPMENT, AND GENETICS OF OPSIN GENE EXPRESSION IN AFRICAN CICHLID FISHES

Kelly E. O'Quin, Doctor of Philosophy, 2011

Directed By: Dr. Karen L. Carleton, Assistant Professor
Department of Biology

The molecular genetic mechanisms that underlie phenotypic evolution include mutations within protein-coding, *cis*-regulatory, and *trans*-regulatory factors. Although many studies have examined how these mutations individually contribute to phenotypic divergence and the formation of new species, none have examined how they may do so collectively. In this study, I examine how these molecular genetic mutations collectively contribute to the evolution of color vision among African cichlid fishes. I show that phenotypic divergence in cichlid color vision is achieved by mutations affecting the coding sequence and expression of seven opsin genes. After contrasting the roles of these two mechanisms, I use bioinformatic-, association-, and experimental genetic analyses to determine what role mutations in *cis*- and *trans*-regulatory DNA play in the evolution of cichlid opsin expression. Specifically, I demonstrate that:

(1) Protein-coding mutations primarily affect cichlid opsins sensitive to the ends of the visible light spectrum (*SWS1* [ultraviolet-sensitive] and *LWS* [red-sensitive]).

(2) Changes in opsin gene expression contribute to large differences in color vision among closely related species. These analyses also reveal that the expression of the *SWS1* and *SWS2B* opsins have diverged among closely related cichlids in association with foraging preferences and ambient light intensity, suggesting that their expression has evolved due to natural selection. Ancestral state reconstructions reveal that changes in opsin expression have evolved repeatedly among cichlids in Lakes Tanganyika and Malawi; further, I find that this repeated evolution has likely been achieved by repeated changes to cichlid development.

(3) Bioinformatic analyses suggest that cichlids have diverged in multiple *cis*-regulatory sequences surrounding the opsin genes, and association mapping identified three putative single nucleotide polymorphisms upstream of the *SWS2A* (blue), *RH2B* (blue-green), and *LWS* (red) opsins that may contribute to cichlid opsin expression differences in *cis*.

(4) Genetic mapping in experimental crosses suggests that divergence in multiple *trans*-regulatory factors also contribute to the evolution of *SWS2B* (violet), *RH2A* (green), and *LWS* (red) opsin expression. The contribution of these *trans*-regulatory factors to the evolution of cichlid opsin expression may outweigh those in *cis*.

These results reveal that multiple molecular genetic mechanisms can contribute to phenotypic evolution among closely related species.

EVOLUTION, DEVELOPMENT, AND GENETICS OF OPSIN GENE EXPRESSION IN AFRICAN CICHLID FISHES

By

Kelly E. O'Quin

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
2011

Advisory Committee:
Assistant Professor Karen L. Carleton, Chair
Associate Professor Michael P. Cummings
Professor Charles Delwiche
Associate Professor Eric Haag
Professor Thomas D. Kocher

Preface

"To suppose that the eye, with all it inimitable contrivances for adjusting the focus to different distances, for admitting different amounts of light, and for the correction of spherical and chromatic aberration, could have been formed by natural selection, seems, I freely confess, absurd in the highest possible degree. Yet reason tells me, that if numerous gradations from a perfect and complex eye to one very imperfect and simple, each grade being useful to its possessor, can be shown to exist; if further, the eyes does vary ever so slightly, and the variations be inherited, which is certainly the case; and if any variation or modification in the organ be ever useful to an animal under changing conditions of life, then the difficulty of believing that a perfect and complex eye could be formed by natural selection, though insuperable by our imagination, can hardly be considered real. How a nerve comes sensitive to light, hardly concerns us more than how life itself first originated . . ."

Charles Darwin, *The Origin of Species*

Dedication

To my family, both genetic and otherwise: thank you for your love and support.

Claire O'Quin

Karen Carleton

Jerry, Linda, Casey, and Katie O'Quin

Michael, Dawn, and Leia Tamplain

Acknowledgements

I would never have undertaken this dissertation without the encouragment of my tenth grade biology teacher, 'Doc' Nathaniel Johnson; I would never have survived the rigors of classwork, teaching, and research without the support and guidance of my advisor, Karen Carleton; and, I would never have finished this work without the sustaining love and patience of my wife, Claire O'Quin. Additionally, I thank each of my committee members—Michael Cummings, Chuck Delwich, Eric Haag, and Tom Kocher—for their time, patience, and insights. Finally, I would also like to thank the following individuals, institutions, and funding sources that supported the research contained in each chapter:

Chapter 2. This chapter was previously published in the journal *PLoS Biology* and I reprint it here in slightly modified form under the Creative Commons Attribution 2.5 Generic License. I would like to thank my co-authors for their patience and hard work on this chapter. Together, we also thank Thomas Kocher, Reade Roberts, Pat Danley, Darrin Hulsey, Todd Streelman, Aimee Howe, and Richard Zatha for their help in collecting Lake Malawi cichlids for this study; and Mhoja Kayeba, Mohamed Haluna, John Mrosso, and Martine Maan for their help in collecting Lake Victoria cichlids. I and the other co-authors would also like to thank the University of Malawi, especially Aggrey Ambali, Lawrence Malekano, and the Tanzania Fisheries Research Institute, especially Professor Philip Bwathondi and Egid F. Katunzi. The cichlid lab group at the University of Maryland provided valuable comments on many earlier drafts of this chapter.

Chapter 3. This chapter was previously published in the journal *Molecular Biology and Evolution* and is reprinted here in slightly modified form under Oxford University Press license no. 2603780800339. I once again thank my co-authors for their hard work and generous help on this chapter, and also want to thank our many funding sources. While working on this chapter, I was supported by awards from the Guy Jordan Endowment Fund of the American Cichlid Association, the Cosmos Club Foundation of Washington, DC, and a Grant-In-Aid of Research from Sigma Xi, The Scientific Research Society. KLC was supported by the University of Maryland and National Science Foundation (NSF) grant IOS-0841270. HAH was supported by NSF grant IOS-021795. We thank S. Renn, C. Shumway, A. Pollen, G. and W. Kazumbe, and S. Bahan for assistance in the field; the Tanzania Fisheries Research Institute, the Tanzania Commission on Science and Technology, and Prof. A. Nikundiwe (University of Dar Es Salaam) for their kind support of our research; Z. Nasser for help sequencing the opsin genes; and D. Soares for providing *A. burtoni* tissues for this study. This work benefited greatly from comments by two anonymous reviewers.

Chapter 4. This chapter was previously published in the journal *Evolution & Development* and is reprinted here in slightly modified form under John Wiley and Sons license no. 2631370564407. I thank my co-authors as well as Christopher Hofmann, Jane Schulte, and other members of the Carleton lab for their comments on earlier versions of this chapter; we also thank two anonymous reviewers for their

helpful comments and suggestions. I was supported by an Ann G. Wylie dissertation fellowship as well as a doctoral fellowship from Drs. Wayne T. and Mary T. Hockmeyer (University of Maryland). KLC was supported by the University of Maryland and National Science Foundation (NSF) grant IOS-0841270.

Chapter 5. This chapter has been accepted for publication, pending minor revision, in the journal *BMC Evolutionary Biology*. I thank my co-authors—many of whom are undergraduate volunteers—for their hard work. Together, we thank Takayuki Katagiri for making the *Oreochromis niloticus* BAC clone library and Bo Young Lee for pooling this library for PCR screening. We also thank Frederica DiPalma for generating the *Meteriaclima zebra* library and Celeste Kidd for screening this library for the opsin-containing BACs. This work was supported with grants to KLC from NSF (IOS-0841270), NIH (R15 EY016721-01) and the University of Maryland. KEO was supported by a Wayne T. and Mary T. Hockmeyer Doctoral Fellowship and an Ann G. Wylie Dissertation Fellowship from the University of Maryland.

Chapter 6. I once again thank my co-authors, the majority of whom are undergraduate volunteers ("the gene team"). Togther, we thank Jen Ser for arranging the genetic cross between *Aulonocara baenschi* and *Tramitichromis intermedius*, and Christopher Hofmann for collecting broods and retinas. This work was supported with grants to KLC from NSF (IOS-0841270), NIH (R15 EY016721-01), and the University of Maryland. KEO was supported by an Ann G. Wylie Dissertation Fellowship from the University of Maryland.

Table of Contents

Preface ii
Dedication iii
Acknowledgements iv
Table of Contents vii
List of Tables ix
List of Figures xii

Chapter 1: The Locus of Evolution: Introduction and Overview to Dissertation 1
Evolutionary Genetics and the "Locus of Evolution" 1
The Vertebrate Visual System 6
The African Cichlid System 12
Overview of Dissertation 16
Figures 19

Chapter 2: The Eyes Have It: Regulatory and Structural Changes Both Underlie Cichlid Visual Pigment Diversity 23
Abstract 24
Introduction 25
Materials and Methods 28
Results 38
Discussion 46
Figures 53

Chapter 3: Parallel Evolution of Opsin Gene Expression in African Cichlid Fishes . 59
Abstract 60
Introduction 61
Materials and Methods 66
Results and Discussion 76
Tables 96
Figures 100

Chapter 4: New Evidence for the Role of Heterochrony in the Repeated Evolution of Cichlid Opsin Expression 105
Abstract 106
Introduction 107
Materials and Methods 112
Results 116
Discussion 121
Conclusions 128
Tables 130
Figures 131

Chapter 5: Divergence in *cis*-regulatory sequences surrounding the opsin gene arrays of African cichlid fishes 134
Abstract 135
Introduction 136
Materials and Methods 142
Results and Discussion 150
Conclusions 179
Tables 181
Figures 188

Chapter 6: Quantitative Genetic Analysis of *trans*-Regulatory Factors Associated with Opsin Gene Expression in African Cichlids 193
Abstract 194
Introduction 195
Materials and Methods 199
Results 203
Discussion 207
Conclusions 212
Tables 214
Figures 218

Chapter 7: The Locus of Evolution from a Cichlids' Eye View: Summary and Conclusion to Dissertation 220
Overview of Dissertation Results 221
Synthesis 223
Future Directions 227

Appendices 233
Appendix 1: Supplementary Tables and Figures for Chapter 2 233
Appendix 2: Supplementary Tables and Figures for Chapter 3 251
Appendix 3: Supplementary Tables and Figures for Chapter 4 266
Appendix 4: Supplementary Tables and Figures for Chapter 5 269
Bibliography 307

List of Tables

Table 3-1. Lake Tanganyika cichlid species used in Chapter 2..... 96

Table 3-2. Sequence of all primers and probes used to measure cichlid opsin gene expression. 98

Table 3-3. Results of phylogenetic ANOVA comparing opsin gene expression with foraging preference and *post hoc* comparisons of *SWS1* expression between foraging levels. 99

Table 4-1. Statistical significance of linear and nonlinear coefficients estimated for *A. burtoni* and *M. auratus* predicted photoreceptor sensitivities. 130

Table 5-1. List of candidate transcription factors surveyed in this study. 181

Table 5-2. Assembly statistics for *O. niloticus* and *M. zebra* opsin-containing BACs. 182

Table 5-3. Comparison of sequence similarity and TFBS/miRNA target site divergence for three putative *cis*-regulatory regions surrounding the opsin arrays of *O. niloticus* and *M. zebra*. 183

Table 5-4. Conserved microRNA target sites within the 3'-UTRs of each opsin in *O. niloticus* and *M. zebra*. 185

Table 5-5. Polymorphism statistics for 8 candidate *cis*-regulatory regions in 18 Lake Malawi cichlid species. 186

Table 5-6. Results of allelic associated between SNPs underlying peaks of nuclotide diversity and opsin expression in 18 Lake Malawi cichlid species. 187

Table 6-1. Candidate *trans*-regulatory factors examined in Chapter 6. 214

Table 6-2. Primers used to amplify and genotype polymorphisms in Chapter 6..... 215

Table 6-3. Results of ANOVA comparing mean single- and double-cone sensitivity among genotypic classes at 15 candidate *trans*-regulatory loci. 216

Table 6-4. Results of ANOVA comparing mean *RH2B*, *RH2A*, and *LWS* opsin expression among genotypic classes at four candidate *trans*-regulatory loci. 217

Supplementary Table S2-1. Sample size, relative opsin gene expression, estimated single- (SC) and double- (DC) cone sensitivity (λ_{max}), visual palette grouping (k = 3

clustering), foraging and habitat preference, and clade membership of all Lake Victoria and Lake Malawi cichlids used in Chapter 2. 234

Supplementary Table S2-2. Accession numbers for all opsins used in the opsin sequence diversity analysis from Chapter 2. 238

Supplementary Table S2-3. Results of phylogenetic ANOVA comparing relative opsin expression and single and double cone sensitivity to foraging and habitat preference among cichlids from Lake Malawi using three phylogenetic hypotheses. 240

Supplementary Table S2-4. Location, depth, Secchi disc readings, *SWS2B* quantum catch (QC), and *SWS2B* relative opsin expression for Lake Victoria taxa from Chapter 2. 241

Supplementary Table S2-5. Summary of amino acid variation in opsin genes of cichlids from Lakes Malawi (LM) and Victoria (LV) presented in Chapter 2. 242

Supplementary Table S2-6. Substitutions between amino acids with different physical properties that are located within the transmembrane regions of the cichlid opsins. 243

Supplementary Table S2-7. Accession numbers for mtDNA sequences used to generate phylogenies for the comparative methods in Chapter 2. 245

Supplementary Table S2-8. Three distance-based validation statistics for the clusters of cichlid opsin expression and photoreceptor sensitivities presented in Chapter 2. 247

Supplementary Table S3-1. Primers used for the detection and sequencing of opsin and mtDNA in cichlids from Lake Tanganyika. 251

Supplementary Table S3-2. Unique primer and probe combinations used to measure opsin gene expression in cichlids from Lake Tanganyika. 253

Supplementary Table S3-3. Relative opsin gene expression (% total ± s.e.), estimated single- (SC) and double- (DC) cone photoreceptor sensitivity (λ_{max}), and lens transmittance (T50) results for 28 cichlids species from Lake Tanganyika. 256

Supplementary Table S3-4. Protein-coding sequence variation in ospisn genes of cichlids from Lake Tanganyika and *Oreochromis niloticus*. 258

Supplementary Table S3-5. Results of Bayesian cluster analysis (k = 3 – 5 clusters) of opsin gene expression from 93 cichlids species from Lakes Tanganyika (LT), Malawi (LM), and Victoria (LV). 261

Supplementary Table S4-1. Relative opsin gene expression (% of total opsin expression) and predicted photoreceptor sensitivity (PS_{max}, nm) of single- (SC) and double- (DC) cones measured for all samples reported in Chapter 4, including *Oreochromis niloticus*..... 266

Supplementary Table S5-1. GenBank accession numbers for all Lake Malawi cichlid sequences generated in Chapter 5. 270

Supplementary Table S5-2. Primers used to amplify and sequence the proximal promter regions and 3'-UTRs of several opsins from 18 Lake Malawi cichlid species. 271

Supplementary Table S5-3. Names, relative opsin expression results, and polymorphisms found with the proximal promoter and 3'-UTR of five opsins in 18 Lake Malawi cichlid species. 272

Supplementary Table S5-4. Complete transcription factor binding site profiles for 23 conserved non-coding elements (CNEs) in *O. niloticus* and *M. zebra*.. 280

Supplementary Table S5-5. Complete list of microRNA target sites identified with the 3'-UTRs of each opsin in *O. niloticus* and *M. zebra*. 282

Supplementary Table S5-6. Length and pairwise sequence divergence (D_{xy}) scores between *O. niloticus* and *M. zebra* for each coding and non-coding region examined. 284

Supplementary File 5-1. FASTA-formatted text file of all 23 conserved non-coding elements (CNEs), opsin proximal promoter and 3'-UTR from *Oreochromis niloticus* and *Metriaclima zebra*, as well as randomly chosen *O. niloticus* non-opsin sequences from Chapter 5. 295

List of Figures

Figure 1-1. Cartoon illustrating the different molecular genetic bases to phenotypic evolution. 19

Figure 1-2. Overview of the vertebrate eye, photoreceptors, and opsin genes. 20

Figure 1-3. Diversity of African cichlid fishes. 21

Figure 1-4. Variation in photoreceptor sensitivity and opsin gene expression in three African cichlids from Lake Malawi. 22

Figure 2-1. Opsin gene expression from all species surveyed. 53

Figure 2-2. Gene expression profiles and single- and double-cone sensitivities form three clusters. 54

Figure 2-3. Selective pressures drive opsin expression with each lake. 55

Figure 2-4. Visual pigment performance in Lake Victoria. 56

Figure 2-5. Visual pigment performance in Lake Malawi. 57

Figure 2-6. The shortest- and longest-wavelength opsins have the greatest sequence diversity. 58

Figure 3-1. Schematic of the East African Great Lakes and the phylogenetic structure of their associated cichlid species flocks. 100

Figure 3-2. Opsin expression diversity in 28 cichlid species from LT. 101

Figure 3-3. Parallel evolution of opsin gene expression in 47 African cichlid fishes from LT, LM, and LV, as well as the rivers (R). 103

Figure 3-4. Comparative analysis of opsin gene expression with foraging preference and lens transmittance. 104

Figure 4-1. Evolution of opsin expression in African cichlids. 131

Figure 4-2. Ontogenetic variation in opsin gene expression for two African cichlids, (A) *Astatotilapia burtoni* and (B) *Melanochromis auratus*. 132

Figure 4-3. Ontogenetic variation in the predicted maximal sensitivity of single- and double-cone photoreceptors for (A) *Astatotilapia burtoni* and (B) *Melanochromis auratus*. 133

Figure 5-1. Pairwise comparison of sequence conservation between *O. niloticus* opsin-containing BAC regions and four fish genomes. 188

Figure 5-2. Alignment of two putative opsin regulatory elements (CNE 10a and b) in fishes. 189

Figure 5-3. Transcription factor binding site diversity within opsin proximal promoters. 190

Figure 5-4. Interspecific polymorphism in eight putative *cis*-regulatory regions from 18 Lake Malawi cichlid species. 191

Figure 5-5. Divergence among coding and non-coding regions in *O. niloticus* and *M. zebra* opsin-containing BAC sequences. 192

Figure 6-1. Variation in opsin gene expression in F_2 hybrid intercross progeny of a genetic cross used to study cichlid opsin expression. 218

Figure 6-2. Expression QTL for opsin gene expression linked two *trans*-regulatory factors. 219

Supplementary Figure S2-1. Depiction of Lake Malawi and Lake Victoria cichlid opsin expression in a phylogenetic context. 248

Supplementary Figure S2-2. Phylogenies used for phylogenetically-control statistical methods (ANOVA and independent contrasts) in Chapter 2. 249

Supplementary Figure S2-3. Synonymous substitution rates (D_S) and nucleotide diversity (*pi*) of each opsin gene for both Lake Malawi and Lake Victoria cichlids. 250

Supplementary Figure S3-1. Parallel evolution of opsin gene expression in 47 African cichlids fishes from Lakes Tanganyika (LT), Malawi (LM) and Victoria (LV), as well as the rivers (R), continued. 264

Supplementary Figure S5-1. Identification of opsin-containing BACs from Finger Printed Contigs (FPCs). 286

Supplementary Figure S5-2. Synteny (Pip plots) of *O. niloticus* opsin-containing BACs with the genome assemblies of five teleost species. 289

Supplementary Figure S5-3. Synteny (Pip plots) of *O. niloticus* and *M. zebra* opsin-containing BAC sequences. 292

Supplementary Figure S5-4. Orthology of *RH2* and *SWS2* opsin paralogs from five teleost fish genomes. 293

Supplementary Figure S5-5. Opsin gene content of five teleost genomes. 294

Chapter 1:
The Locus of Evolution: Introduction and Overview to Dissertation

Evolutionary Genetics and the "Locus of Evolution"

Evolutionary Genetics: Past and Present

In the 150 years since Darwin published his seminal treatise *On The Origin of Species* (Darwin 1859), our understanding of biological evolution has advanced by leaps and bounds. In *The Origin*, Darwin synthesized evidence from both natural and artificial systems to support the hypothesis that biological evolution results from selection among the individuals of a population undergoing the "struggle for life" on the basis of heritable variation in their traits (Darwin 1859). Some 70 years later, this theory of evolution via natural selection was assimilated with the modern theory of genetics, and expanded to include other mechanisms of evolutionary change, including genetic drift, mutation, and migration (Fisher 1930; Dobzhansky 1937; Huxley 1942). With this Modern Synthesis came a rigorous mathematical framework for examining evolution within populations, leading to the emergence of the field of evolutionary genetics (Dietrich 2006). At its heart, this broad field aims to understand the evolutionary and genetic mechanisms of phenotypic and molecular divergence in natural populations, taking genetics "from the laboratory to the field" (Dietrich 2006; Fox and Wolf 2006). Advances in the field of molecular biology further revolutionized evolutionary genetics, allowing biologists to examine genetic

variation at numerous levels of biological organization, including whole genomes (Lewontin 1991; Graur and Li 2000; Fox and Wolf 2006). In the past 30 years alone, evolutionary geneticists have shined new light on our understanding of evolution by connecting divergence in specific organismal phenotypes with variation at specific chromosomal regions, genes, and even nucleotides (for a general overview of some of these traits and their underlying loci, see Stern (2006; 2007; 2008; 2008). Now with the advent of affordable DNA sequencing and large repositories of genetic and phenotypic data, the field of evolutionary genetics is poised to address many long-standing problems in evolutionary biology, including analysis of the molecular genetic mechanisms of phenotypic divergence—the "locus of evolution" (Hoekstra and Coyne 2007).

The Molecular Genetic Basis of Phenotypic Evolution

The molecular genetic mechanisms underlying functional diversification and phenotypic evolution can be divided into a few major categories (Figure 1-1). First, mutations that arise within the protein-coding region of genes can alter the structure and function of these genes and their products, thereby altering organismal phenotypes. For much of the last 70 years since the Modern Synthesis, this class of mutations has formed the basis of most genetic models of phenotypic evolution. For example, two cases of vertebrate phenotypes that have evolved due to mutations affecting the protein-coding region of genes include (*i*) resistance to the drug Warfarin among humans (*Homo sapiens*) and rats (*Rattus norvegicus*) due to missense mutations within the vitamin K epoxide reductase gene *VKORC1* (Rost et al. 2004) and (*ii*) cryptic melanism among beach mice (*Peromyscus polionotus*) due

to a single amino acid substitution within the melanocortin receptor gene *Mc1r* (Hoekstra et al. 2006). In both cases, a small number of mutations in the protein-coding region of each gene have been associated with changes in that gene's activity or function, and result in phenotypic changes mirroring those seen in natural populations. These and many other examples (see Table 1 in Hoekstra and Coyne (2007)) provide powerful evidence that evolutionary change in organismal phenotypes can result from mutations within the protein-coding regions of genes.

However, as long as 40 years ago, before many links could be made between specific mutations and their phenotypic consequences, several biologists had posited that mutations affecting non-coding regulatory DNA could also drive the evolution of organismal phenotypes (Britten and Davidson 1971; King and Wilson 1975). Instead of altering protein structure and function, mutations in non-coding DNA alter organismal phenotypes by changing the type, location, timing, or amount of protein expressed. One of the earliest arguments for the role of these regulatory mutations in phenotypic evolution came from comparisons of protein sequence divergence between chimps (*Pan troglodytes*) and humans (King and Wilson 1975). King and Wilson (1975) summarized early studies of protein sequence variation in these two species and found that they were divergent in < 1% of their protein sequences. Consequently, King and Wilson (1975) argued that mutations in protein-coding regions are insufficient to account for the vast morphological, physiological, and behavioral differences observed between humans and chimps. They argued instead that changes in anatomy and way of life are more often the result of mutations that affect the expression of genes rather than the genes themselves. Further, protein-

coding genes comprise < 5% of the human genome (Lander et al. 2001), revealing a potentially large source of regulatory DNA; however, examples of specific regulatory mutations that alter gene expression and organismal phenotypes have been scarce, especially within vertebrate systems. This dearth of examples is due in part to the infancy of comparative gene expression (Wray 2007) as well as the inherent difficulty in identifying and testing the significance of variation in such regulatory sequences outside of a well-understood genetic code (Chen and Stephan 2006). But two recent examples of animal phenotypes that have evolved due to mutations affecting non-coding regulatory DNA include: (*i*) lactase persistence in humans due to several mutations within a *cis*-regulatory region found ~ 22 kb upstream of the lactase gene *LCT* (Enattah et al. 2002; Olds and Sibley 2003; Tishkoff et al. 2007) and (*ii*) wing pigmentation in *Drosophila biamripes* due to mutations within a *cis*-regulatory element found ~ 1 kb upstream of the pigmentation gene *yellow* (Gompel et al. 2004). In both cases, expression of the mutated region in cell culture or in vivo within other *Drosophila* species, respectively, recapitulated the phenotype and gene activity seen in natural populations. Thus, there is growing and evidence (see Table 1 in Wray (2007)) that mutations within non-coding regulatory DNA can also drive phenotypic divergence and evolution.

Most recent studies of mutations within regulatory sequences have focused on mutations that occur *cis* (in linkage) to the genes they regulate. The reasons for this are three-fold. First, since they are found relatively close (generally 1 – 20 kb upstream) to the genes they regulate, *cis*-regulatory sequences represent a natural and tractable first choice in the search for regulatory mutations (Wray 2007). Second, the

promoters of many genes contain multiple *cis*-regulatory sequences that direct gene expression within specific tissues and developmental stages (Carroll 2008). The types of genes with such 'modular' promoter elements are geneally widely-expressed developmental regulatory genes (i.e., signaling pathway components and transcription factors). The wide spatial and temporal expression of these genes means that any mutations within their coding sequences will potentially alter numerous phenotypes, a phenomenon termed mosaic pleiotropy. However, a mutation within in any one *cis*-regulatory sequence will only change gene expression in a few tissues or contexts. Thus, the modularity of *cis*-regulatory elements facilitates the fine-tuning of gene expression and potentially limits much of the negative pleiotropy that would result from protein-coding mutations (Stern 2000; Hittinger et al. 2005; Carroll 2008). Finally, *cis*-regulatory alleles on different chromosomes are transcribed independently, making them co-dominant. This pattern of inheritance is in contrast to many protein-coding mutations, which are generally recessive (Graur and Li 2000). The ability to produce a distinct expression phenotype in heterozygotes makes *cis*-regulatory mutations efficient targets for natural selection (Hartl and Clark 2006). These latter two features of *cis*-regulatory sequences are so appealing that many evolutionary developmental biologists have posited that the evolution of *cis*-regulatory sequences has played a major role in the evolution of animal form (Stern 2000; Carroll 2005; Wray 2007).

Importantly, however, mutations that affect gene expression can also occur *trans* (out of linkage) to the genes they regulate (Jones et al. 1988; Wittkopp et al. 2004). For example, protein-coding mutations within transcription factors can act as

regulatory mutations for potentially hundreds of genes (Carroll 2005, 2008), while mutations within non-coding RNAs (microRNAs) can change levels of gene expression post-transcriptionally (He and Hannon 2004; Chen and Rajewsky 2007) (Figure 1-1). Other mechanisms of phenotypic evolution, such as the post-translational modification of proteins, or epigenetic regulation of genes and phenotypes, are relatively poorly understood and will not be covered here (though this is not to imply that these mechanisms are unimportant for phenotypic evolution). Thus, although strong arguments and evidence exist that regulatory mutations can alter organismal phenotypes, many questions remain. For example, are some phenotypes more likely to evolve by protein-coding mutations than regulatory mutations? Do protein-coding and regulatory mutations differ in their quantitative or qualitative affects on phenotypic evolution? And if regulatory mutations are an important force in driving phenotypic divergence, then where do these mutations typically occur (in *cis* or *trans*) relative to the genes they regulate?

The Vertebrate Visual System

The Utility and Structure of the Vertebrate Visual System

Vision in vertebrates is a complex sensory process that is mediated at several levels of organization, from the anatomical (eye), to the molecular (opsins), to the neuronal (visual cortex). This system is an excellent model with which to study the contributions that different molecular genetic mechanisms make to phenotypic evolution and adaptation since: (*i*) the molecular basis of vision has been studied in numerous vertebrate systems, including humans, for many years and is well

documented (Wald 1968; Nathans et al. 1986; Yokoyama and Yokoyama 1990); (*ii*) phenotypic variation in the visual system of vertebrates is observed at multiple levels of organization, from whole eyes to individual molecules (Yokoyama and Yokoyama 1996; Yokoyama et al. 2008), and (*iii*) the visual system has the potential to profoundly affect organismal fitness through foraging, mate choice, and predation (Lythgoe 1979; Yokoyama and Yokoyama 1996).

At its broadest level of organization, the eye mediates vision in vertebrates by responding to the narrow band of electromagnetic radiation found within the visible light spectrum (360 – 750 nm). But as photons of light strike the eye, they must first pass through the cornea, aqueous humor, crystalline lens, and vitreous humor before they reach the true sensory center of the eye, the retina (Figure 1-2). The retina is a thin sheet of light-sensitive tissue that lines the back of the eye. It is composed of several layers of interconnected nerve cells and receptors and is responsible for many of the intial and intermediate stages of visual processing. Given this role, it is perhaps unsurprising that the retina is actually an outgrowth of the brain and central nervous system (Dowling 1987). During the initial stages of visual perception, photons of light that have passed through the eye then pass through the various nerve layers of the retina before finally striking the light-sensitive photoreceptor cells (Figure 1-2). Once struck, however, these photoreceptor cells quickly set in motion a signal transduction cascade that sends an electrical impulse to the various layers of the retina; it is here that the intermediate phases of visual perception occurs. Numerous nerve cells—including horizontal-, bipolar-, amacrine-, and ganglion-cells—filter, combine, compare, and contrast signals from different photoreceptors

before sending a final electrical impulse down the optic nerve and to the brain (Kolb 2003). Once there, the finer details of visual processing occur within the visual cortex.

Vertebrate Photoreceptors and Spectral Tuning

Although vertebrate vision may be processed at several levels, the retinal photoreceptors have been the chief focus of most studies of vertebrate vision, since it is these cells that are immediately responsible for responding to light. Vertebrates exhibit great variation in the size, shape, and number of their photoreceptor cells, but together they broadly exhibit two photoreceptor types: (*i*) long, rod-shaped cells responsible for dim-light (scotopic) vision, and (*ii*) short, cone-shaped cells responsible for bright-light (photopic [e.g., color]) vision. In some vertebrates, the cone photoreceptors are further subdivided into short, distinct single-cone cells, and longer, joined double-cone cells (Figure 1-2). The light-sensitive property of the photoreceptors is due to a vast array of integral membrane protein-chromophore complexes that cover part of the photoreceptor cell: these protein-chromophore complexes are termed visual pigments (Wald 1968). A visual pigment comprises a vitamin A_1-derived chromophore (11-*cis* retinal) bound to an opsin protein (a seven transmembrane G protein-coupled receptor) (Terakita 2005) (Figure 1-2). When struck by light, the chromophore isomerizes to all-*trans* retinal, causing the opsin to activate the protein transducin, setting off a signal transduction cascade that ultimately results in the passage of an electrical impulse to the several nerve layers of the retina and brain.

Vertebrates vary in the wavelength of light that is maximally absorbed by their photoreceptors and visual pigments (termed λ_{max}); this variation in turn determines the colors and visual signals that different species can perceive and respond to (Yokoyama 2008). Visual sensitivites can also be altered through changes to the ocular media (cornea, aqueous humor, crystalline lens, and vitreous humor) or visual pigment. In the case of the ocular media, the transmissive properties of the crystalline lens and oil droplets can be used to limit or shift the available wavelengths of light before they reach the retina (Boettner and Wolter 1962; Siebeck and Marshall 2001). The transmissive properties of lenses vary among many fish species (Lythgoe 1979; Siebeck and Marshall 2001), while the use of oil droplets is common among birds and reptiles (Bowmaker et al. 1997). However alterations to the opsin protein and retinal chromophore can also change the λ_{max} of visual pigment and, thus, visual sensitivities (Wald 1968). In some vertebrates, replacing the vitamin A_1-derived chromophore 11-*cis* retinal with the vitamin A_2-derived chromophore 11-*cis*-3, 4 didehydroretinal alters the λ_{max} of the visual pigments. 11-*cis*-3, 4 didehydroretinal has one more double-bond than 11-*cis* retinal and, all else being equal, will cause the visual pigment to absorb light of longer wavelengths (Wald 1968). Many fish species use chromophore shifts to tune the λ_{max} of their visual pigments in response to longer-wavelength light present in winter months (Lythgoe 1979). Additionally, vertebrate vision can also be tuned by altering the amino acid sequence of the opsin protein, particularly at residues that surround the chromophore (termed the retinal binding pocket). Numerous studies in a wide array of vertebrates have demonstrated that the λ_{max} of vertebrate visual pigments are tuned to their local light environment by amino

acid substitutions in the opsin proteins, despite the fact that it is feasible to tune sensitivity at higher orders of organization as well (for example visual processing in the retina or brain) (Bowmaker 1995; Yokoyama and Yokoyama 1996; Yokoyama et al. 1999; Hunt et al. 2001; Terai et al. 2002; Carleton et al. 2005a; Sugawara et al. 2005; Terai et al. 2006; Seehausen et al. 2008). This observation suggests that the opsins play an important role in the evolution of vertebrate visual systems. Consequently, the opsins of vertebrates have become a classic example of how protein-coding mutations can influence sensory adaption and phenotypic evolution (Yokoyama 2008).

Opsin Diversity and Fishes

Vertebrates have four classes of opsin proteins, each generally sensitive to a different portion of the visible light spectrum (Yokoyama and Yokoyama 1996). *RH1* or *rhodopsin* mediates scoptic (dim-light) vision and is generally sensitive to green light (490 – 570 nm). *RH2* or *rhodopsin-like* opsins help mediate photopic (color) vision and are also sensitive green light. *SWS* or *short-wavelength-sensitive* opsins also mediate color vision but typically absorb ultraviolet to blue light (360 – 470 nm). Finally, *MWS/LWS* or *middle/long-wavelength-sensitive* opsins also contribute to photopic vision and are sensitive to green and red light (500 – 570 nm). Although the ancestral vertebrate lineage possessed opsins of all four classes, many opsins have been gained or lost independently in different vertebrate lineages (Yokoyama and Yokoyama 1996). For example, most mammals have only three opsins: *RH1*, *SWS*, and *MWS*. Since they have only two opsins used for photopic vision, these species are dichromats and do not possess true color vision. However,

several primates, including humans, have gained an additional opsin through duplication and subsequent protein-coding and regulatory evolution of the *MWS* opsin. This gain has resulted in the evolution of a fourth functional opsin (*LWS*) in primates that is sensitive to red light and which confers full trichromacy, or color vision (Jacobs 1996).

Among vertebrates, aquatic taxa exhibit some of the most diverse visual systems—this is especially true of fishes. The astounding visual diversity of aquatic vertebrates is in part due to two factors. First, the spectral bandwidth of visible light available in aquatic environments varies much more dramatically than in terrestrial environments, since water attenuates light from the short- and long-wavelength regions of the VLS with increasing depth and turbidity (Levine and MacNichol Jr. 1979; Yokoyama 2008). Consequently, many closely-related fish species vary in the λ_{max} of their photoreceptors and visual pigments in response to the depth and light environment of their preferred habitat. Perhaps the clearest example of this diversity is seen in the *RH1* and *SWS* opsin sequences of cottoid fishes (sculpins) from Lake Baikal. Lake Baikal is the deepest freshwater lake in the world, and it is inhabited by several species of sculpins that are segregated by depth. The rod and cone photoreceptors of surface-dwelling (littorral) species exhibit λ_{max} that are long-wavelength-shifted relative to deep-dwelling (abyssal) species. These changes in λ_{max} are reflected in specific amino acid replacements within the *RH1* and *SWS* opsins, and parallel the change in the spectrum of visible light available at greater depths (Bowmaker et al. 1994; Hunt et al. 1996; Cowing et al. 2002). Second, many fish lineages have acquired multiple opsin genes through several bouts of localized and

whole-genome duplication. Thus, although the ancestral vertebrate lineage had four opsins, and most terrestrial mammals have three or four, many fish species have five or more genes covering each opsin class. For example, zebrafish (*Danio rerio*), which is a model fish species, has 9 distinct opsins that collectively confer sensitivity to nearly the entire visible light spectrum (Chinen et al. 2003). In this sense, the opsins of fishes represent an ideal system in which to study the evolution of the visual system in vertebrates.

The African Cichlid System

African cichlids (Perciformes: Cichlidae) comprise a group of freshwater, teleost fish found throughout the lakes and rivers of Africa. Although these fishes are distributed throughout the African continent, the largest and best-characterized assemblages of cichlids are found in the East African Great Lakes and their surrounding rivers (Fryer and Iles 1972) (Figure 1-3). Over 1000 cichlid species have evolved in these lakes within the past 10 million years (MY). Furthermore, these species are incredibly diverse, exhibiting a level of phenotypic diversity that is unparalleled in most other families or orders of vertebrates (Kocher 2004; Sturmbauer 2005; Seehausen 2006). Examples of this diversity are seen in numerous phenotypes, including: (*i*) the size, shape and color of their bodies, (*ii*) their social, ecological and sexual behaviors; and (*iii*) several additional traits associated with ecological and sexual adaptation (Figure 1-3). These two features—the large number of phenotypically diverse species (constituting 2 – 5% of all vertebrates on earth) and the short time-frame in which they arose (most within the last 1 – 2 MY alone)—

makes African cichlids one of a few paradigmatic example of adaptive radiation (Fryer and Iles 1972; Greenwood 1974; Streelman and Danley 2003; Kocher 2004; Seehausen 2006), and their evolution "among the most celebrated events in the history of life" (Schluter 2000). But perhaps more importantly, these same two features also make African cichlids an exciting and tractable system for evolutionary genetic studies.

As noted above, the best-characterized assemblage of African cichlids is found in the Great Lakes of Africa, Lakes Tanganyika, Malawi, and Victoria (Figure 1-3). Lake Tanganyika (LT) contains the oldest lineages of Great Lake cichlids: lineages within this lake are 9 – 12 MY old and consist of ~ 250 species from at least 50 genera and twelve distinct clades or 'tribes' (Fryer and Iles 1972; Brichard 1989; Takahashi 2003). While some of these tribes evolved de novo within LT proper in the last 5 MY, others clearly arose long ago in the rivers and invaded the lake independently (Salzburger et al. 2002). Cichlids from LT are phylogenetically and phenotypically the most diverse assemblage of East African cichlids (Huber et al. 1997; Salzburger et al. 2002; Pollen et al. 2007). Lake Victoria (LV) contains the second oldest radiation of Great Lake cichlids: the monophyletic assemblage found in this lake is approximately 3 MY old (Elmer et al. 2009) and consists of > 500 species. In contrast to the assemblage from LT, however, all LV cichlids are derived from a single tribe that arose within Lake Tanganyika 5 MY ago, the Haplochrominii. Finally, Lake Malawi (LM) contains the youngest group of Great Lakes cichlids: this monophyletic assemblage is 1 – 2 MY old (Koblmüller et al. 2008) and consists of > 600 species from approximately 50 genera. Like cichlids from LV, most LM cichlids

are members of the tribe Haplochromini, and the assemblages ("species flocks") of these two lakes are reciprocally monophyletic (Meyer et al. 1990) (Figure 1-3). Additionally, the monophyletic assemblage of LM's cichlids is itself composed of two reciprocally monophyletic clades, the *mbuna* (translated as "rockfish") and non-*mbuna* (sometimes called *utaka*, although some taxonomists restrict this definition to just two genera of non-*mbuna* (Konings 2007; Oliver 2009)). However, despite the extremely close phylogenetic (Meyer et al. 1990; Kocher et al. 1993) and genetic (Loh et al. 2008) affinity of cichlids in LM, these species are phenotypically quite diverse. In fact, many LM cichlids have convergently evolved numerous traits in common with the older and more diverse cichlids of LT (Kocher et al. 1993; Kassam et al. 2003). Thus, the cichlid system is ideal for evolutionary genetic analysis, since sufficient time has passed for phenotypic diversity to arise among its many species, but not so much that the molecular and phenotypic signatures left by the evolutionary forces that produced this diversity have been erased by subsequent diversification and mutation (see Storz (2005)).

The visual system of African cichlids is an especially good model with which to study the roles of protein-coding and regulatory mutations during phenotypic evolution. Cichlids have eight opsin genes, seven used for photopic vision and one used for scotopic vision (Carleton 2009). These opsins are *SWS1* (ultraviolet-sensitive), *SWS2B* (violet-sensitive), *SWS2A* (blue-sensitive), *RH2B* (blue–green-sensitive), *RH2A*α and *RH2A*β (green-sensitive), *LWS* (red-sensitive), and *RH1* (dim-light-sensitive) (Spady et al. 2006). As in other vertebrates (Yokoyama and Yokoyama 1996), protein-coding mutations have played an important role in

evolution of spectral sensitivity in cichlids. For example, among cichlids from LV, polymorphisms in the protein-coding sequence of the *LWS* opsin are adaptively associated with variation in the local light environment and male color, potentially contributing to ecological speciation in these species (Terai et al. 2006; Seehausen et al. 2008). Additionally, virtually all of the cichlid opsins exhibit molecular signatures of adaptive protein-coding evolution (Sugawara et al. 2002; Spady et al. 2005) (but see Yokoyama et al. (2008)). Typically, these protein-coding mutations are associated with changes in spectral sensitivity (λ_{max}) of 1–15 nm (Yokoyama 2008).

However, in addition to protein-coding mutations, cichlids from LM exhibit large, 30 – 100 nm differences the λ_{max} of their photoreceptors that cannot be explained by simple protein-coding mutations within their opsins (Parry et al. 2005; Jordan et al. 2006) (Figure 1-4). This scenario is reminiscent of the contrast between protein-coding and phenotypic divergence seen in chimps and humans (King and Wilson 1975). Regulatory mutations that alter the relative expression of different groups of opsins have resulted in the evolution of many closely related cichlids that have photoreceptors sensitive to very different regions of the visible light spectrum (Figure 1-4). Additionally, among LV cichlids of the genus *Pundamilia*, variation in the ratio of red:green sensitive double cones may represent an adaptation to local differences in the light environment (Carleton et al. 2005a). Since these species have distinct red- and green-sensitive opsins like other African cichlids, this variation in cone sensitivity implies a variation in the ratio of *LWS*:*RH2A* expression. Thus, preliminary studies suggest that regulatory mutations can also contribute to the evolution of visual sensitivity in African cichlids, and therefore vertebrates in general.

But since each of these previous studies (Carleton et al. 2005a; Parry et al. 2005; Jordan et al. 2006) sampled only 3 – 10 species, it is unclear to what extent this is true for the remainder of cichlids in LM and LV, or for African cichlids from other lakes. For in contrast to cichlids from LM and LV, little is known of the visual sensitivities of cichlids from LT, except for one lone species, *Astatotilapia burtoni* (Fernald 1981).

Overview of Dissertation

Despite the initial results of Carleton et al. (2005a), Parry et al. (2005), and Spady et al. (2006), many intriguing questions remain regarding the evolutionary genetics of cichlid visual sensitivity. If regulatory mutations can contribute to the evolution of visual sensitivity among African cichlids, are these mutations adaptive, as has been demonstrated for the many opsin protein-coding mutations? Do regulatory mutations contribute to similar shifts in spectral sensitivity (1 – 15 nm) as other protein-coding mutations, or do these two types of mutations have distinct effects on phenotypic evolution? Do regulatory mutations represent an equally important source of genetic variation as protein-coding mutations, are they more important, or vice versa? And if regulatory mutations indeed play an important role in the evolution of cichlid visual sensitivity, then what and where are these mutations located relative to the opsins, in *cis* or *trans*? These are the questions that this dissertation will attempt to address. In summary the goals of this dissertation are to: (*i*) examine the diversity and evolutionary history of opsin gene expression among African cichlids from Lakes Tanganyika, Malawi, and Victoria; (*ii*) test hypotheses

concerning the evolutionary and developmental mechanisms by which opsin gene expression has evolved in cichlids; and (*iii*) elucidate the genetic factors that underlie variation in cichlid opsin expression. To this end, I present five chapters of research that each examines one or more of these questions.

In Chapters 2 and 3 (Hofmann et al. 2009; O'Quin et al. 2010), we examine levels of opsin gene expression in approximately 100 cichlid species and populations from Lakes Tanganyika, Malawi and Victoria. We then estimate the impact of changes in gene expression on the maximal sensitivity of cichlid single- and double-cone photoreceptors. In order to determine whether any changes in opsin expession may be due to natural selection, we test whether divergence in opsin expression is associated with divergence in several ecological variables. We compare and contrast interspecific variation in opsin expression with interspecific variation in the amino acid sequence of each opsin. Finally, we also examine the evolutionary history of opsin expression among cichlids in these lakes. By comparing regulatory and protein-coding variation in the same opsins and species, we elucidate the unique contributions that both molecular genetic mechanisms make to phenotypic divergence in cichlid visual sensitivity.

In Chapter 4 (O'Quin et al. 2011) we test the hypothesis that the repeated evolution of opsin expression patterns among African cichlids in LT and LM has been facilitated by changes to cichlid development (heterochrony). We examine opsin gene expression in developing fry of two cichlid species (one of them phylogenetically intermediate to cichlids from LT and LM) then compare this to developmental variation observed in their distant ancestor.

In Chapters 5 and 6 (O'Quin et al., accepted; O'Quin et al. unpublished data) we attempt to identify the *cis*- and *trans*-regulatory factors that underlie variation in cichlid opsin expression. In Chapter 5, we examine putative *cis*-regulatory sequences surrounding the opsin arrays of cichlids with different developmental and adult opsin expression patterns. We examined three types of putative *cis*-regulatory sequences—conserved non-coding elements, proximal promoter regions, and 3'-UTRs—and compare them among two cichlid species for evidence of divergence. Finally, we resequence those regions that exhibit significant divergence in an interspecific panel of cichlids from Lake Malawi in order to identify single nucleotide polymorphisms (SNPs) associated with changes in opsin expression in *cis*. In Chapter 6, we examine genetic divergence at several candidate *trans*-regulatory factors within the progeny of an experimental cross between two Lake Malawi cichlids. The candidate *trans*-regulatory factors we examine include 15 transcription factors that have been shown to influence opsin expression in several fish or vertebrate model systems.

Finally, In Chapter 7, I revisit "the locus of evolution" in light of our results regarding the evolution, development, and genetics of cichlid opsin expression—a cichlid's eye view of the locus of evolution. I synthesize information from this dissertation and other studies regarding the roles of protein-coding, *cis*-, and *trans*-regulatory mutations and the evolution of cichlid visual sensitivity. I suggest what the direction of future work should be and make a final statement on the evolutionary genetics of phenotypic evolution.

Figures

Figure 1-1. Cartoon illustrating the different molecular genetic bases to phenotypic evolution. Solid black bars represent protein-coding DNA, thin black lines represent non-coding DNA. Colored star-bursts represent mutations in each DNA sequence, respectively.

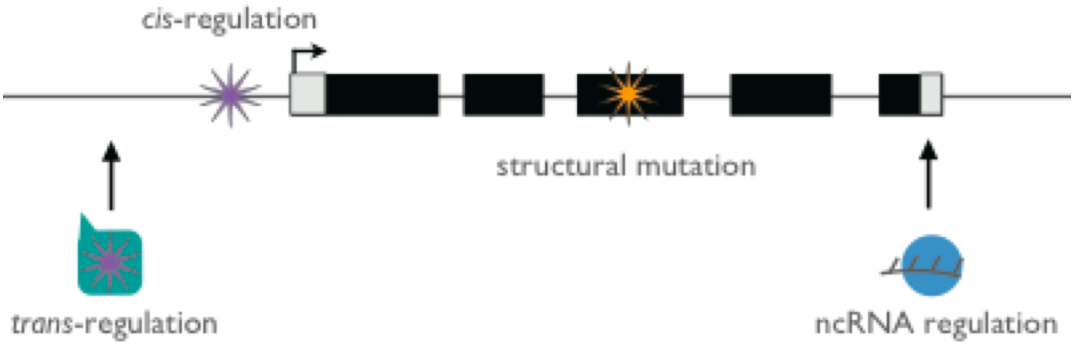


Figure 1-2. Overview of the vertebrate eye, photoreceptors, and opsin genes. (A) Generalized vertebrate eye with the iris, lens, retina, and vitreous humor labeled. (B) Single- and double-cone photoreceptors used for photopic vision and their spatial arrangement in the retina. The color of each photoreceptor corresponds to the wavelengths of light (blue, green, or red) that each is generally sensitive to. The spatial arrangement of four pairs of double-cones surrounding one single-cone is common among fishes and African cichlids in particular. (C) Seven transmembrane structure of a generalized opsin protein. Circles correspond to amino acid residues and gray bars denote transmembrane-spanning regions. The structure of the A_1-derived chromophore, 11-*cis* retinal, is shown in red.

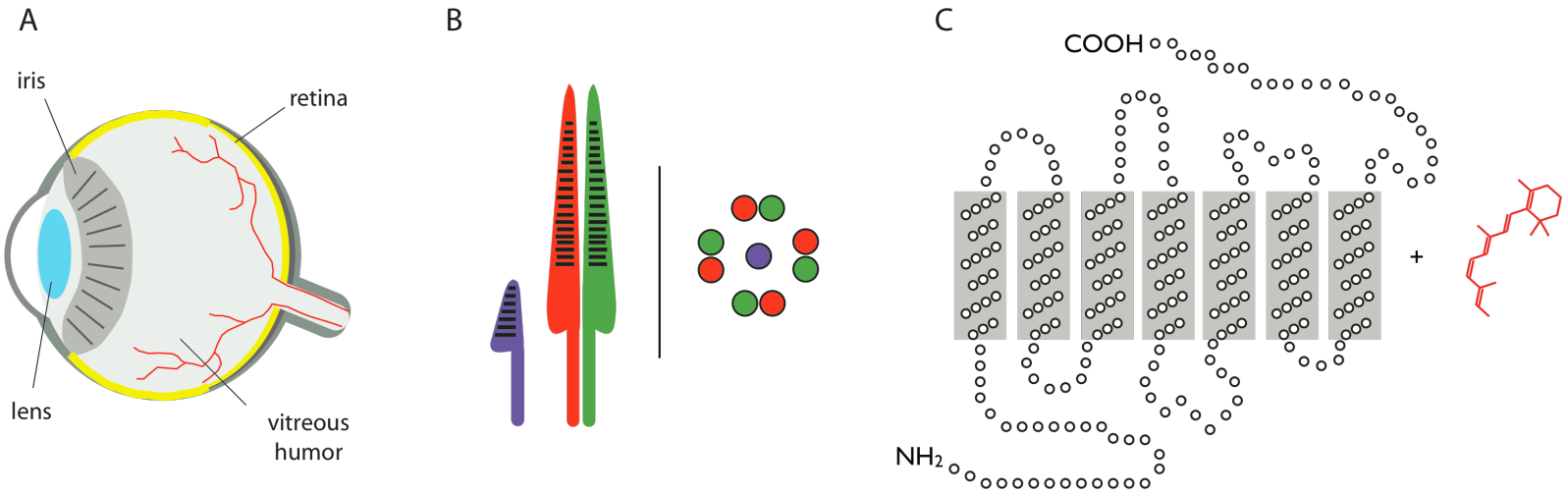
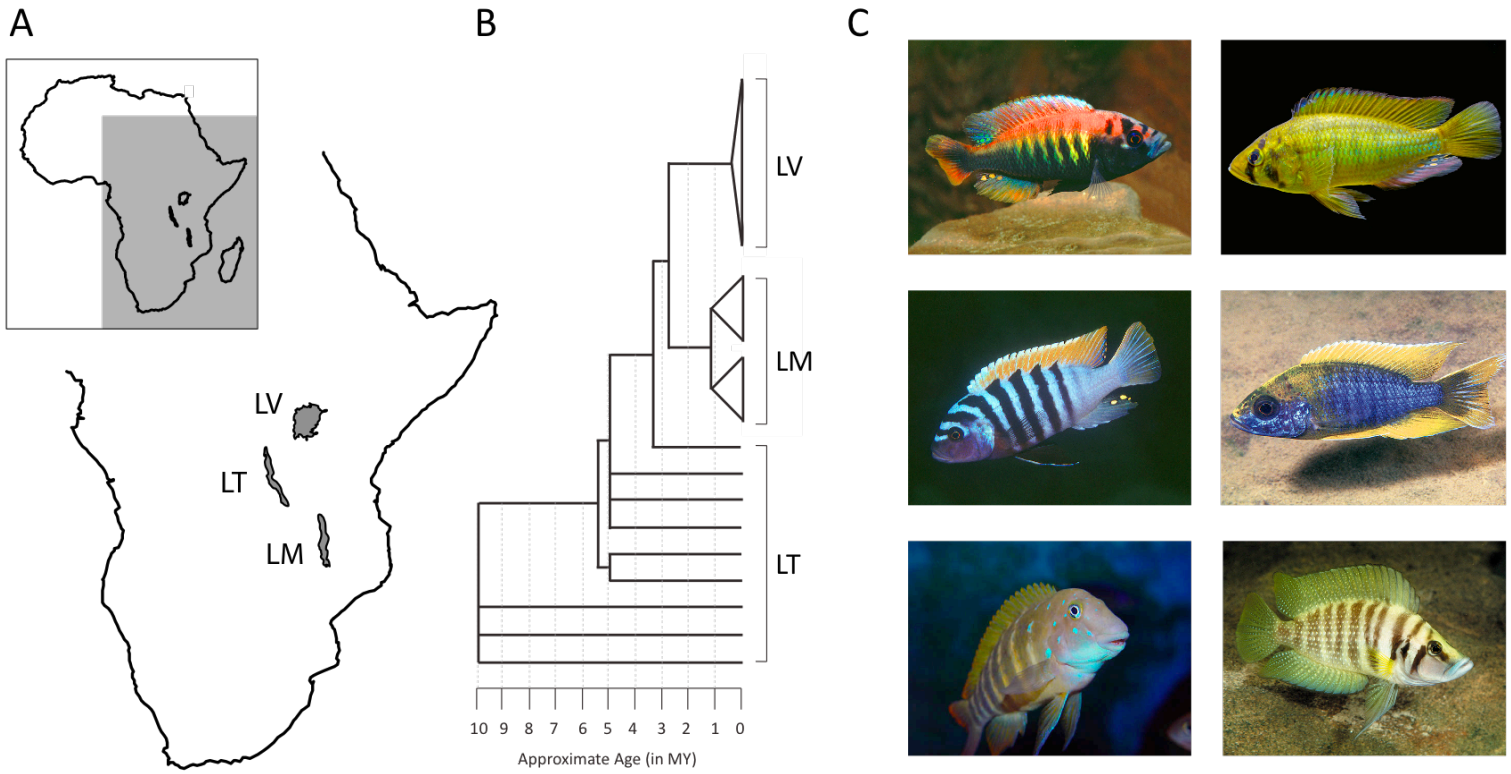
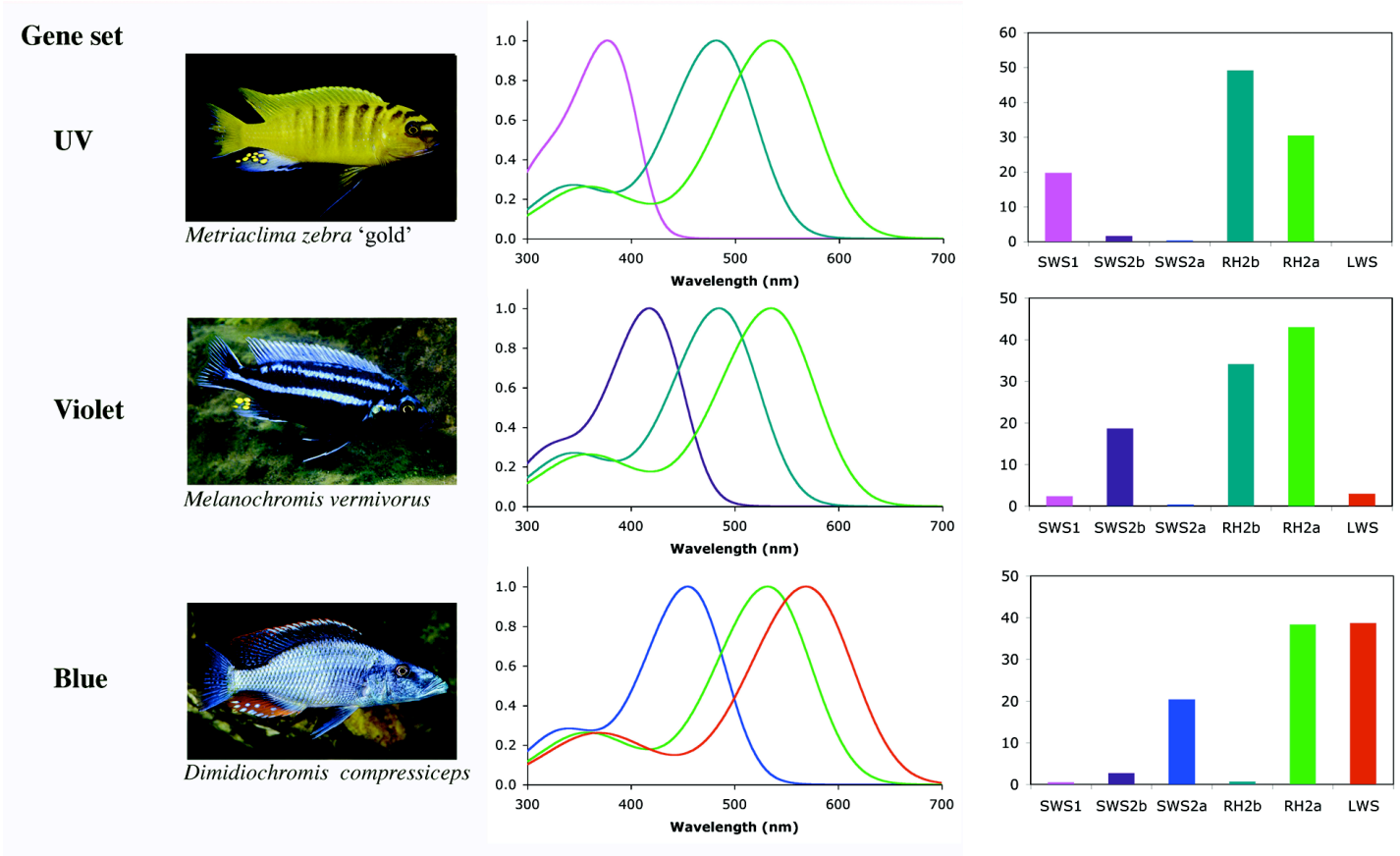


Figure 1-3. Diversity of African cichlid fishes. (A) Map of African continent with the three African Great Lakes—Lakes Tanganyika (LT), Malawi (LM), and Victoria (LV)—highlighted in gray. (B) Broad phylogenetic relationships of cichlids from each lake. (C) Examples of African cichlids and their phenotypic diversity*.



* Species names and photo credits for cichlids in part C, clockwise from top right: *Astatotilapia allaudi* (Kevin Bauman, www.african-cichlid.com); *Aulonocara mamalela* (Ad Konings, www.cichlidpress.com); *Altolamprologus calvus* (Ad Konings); *Eretmodus cyanostictus* (Paolo Salvagiana, www.malawicichlidhomepage.com); *Metriaclima zebra* (Ad Konings); *Pundamilia nyerei* (Ad Konings)

Figure 1-4. Variation in photoreceptor sensitivity and opsin gene expression in three African cichlids from Lake Malawi. First panel illustrates three cichlid species from Lake Malawi; second panel, photoreceptor absorbance of single- and double-cones colored by their wavelength of maximum absorbance (λ_{max}); third panel, relative opsin gene expression. This figure is reproduced with permission from Carleton (2009) (John Wiley and Sons license # 2617691236760).



Chapter 2:

The Eyes Have It: Regulatory and Structural Changes Both Underlie Cichlid Visual Pigment Diversity

Christopher M. Hofmann *, Kelly E. O'Quin*, N. Justin Marshall, Thomas W. Cronin, Ole Seehausen, and Karen L. Carleton

*These authors contributed equally to this work.

See Appendix 1 for all supplementary tables (Tables S2-1 – S2-8) and figures (Figures S2-1 – S2-3) referenced in this chapter.

This chapter is published as:

Hofmann CM, O'Quin KE, Marshall NJ, Cronin TW, Seehausen O, Carleton, KL (2009) The Eyes Have It: Regulatory and Structural Changes Both Underlie Cichlid Visual Pigment Diversity. *PLoS Biology* 7(12): e10000266.

Abstract

A major goal of evolutionary biology is to unravel the molecular genetic mechanisms that underlie functional diversification and adaptation. We investigated how changes in gene regulation and coding sequence contribute to sensory diversification in two replicate radiations of cichlid fishes. In the clear waters of Lake Malawi, differential opsin expression generates diverse visual systems, with sensitivities extending from the ultraviolet to the red regions of the spectrum. These sensitivities fall into three distinct clusters and are correlated with foraging habits. In the turbid waters of Lake Victoria, visual sensitivity is constrained to longer wavelengths, and opsin expression is correlated with ambient light. In addition to regulatory changes, we found that the opsins coding for the shortest- and longest-wavelength visual pigments have elevated numbers of potentially functional substitutions. Thus, we present a model of sensory evolution in which both molecular genetic mechanisms work in concert. Changes in gene expression generate large shifts in visual pigment sensitivity across the collective opsin spectral range, but changes in coding sequence appear to fine-tune visual pigment sensitivity at the short- and long-wavelength ends of this range, where differential opsin expression can no longer extend visual pigment sensitivity.

Introduction

A very large body of literature has been dedicated to the geography, ecology, and genetics of adaptive diversification and speciation (Darwin 1859; Schluter 2000; Coyne and Orr 2004; Gavrilets 2004; Price 2007). Yet, the proximate mechanisms responsible for diversification have been characterized for only a few traits in a few systems (Coyne and Orr 2004). The molecular genetic mechanisms underlying functional diversification can be divided into two major categories. First, changes in gene expression (either through *cis*- or *trans*-acting regulatory factors) can alter the type, location, timing, or amount of protein produced. Alternatively, changes in gene coding sequence can alter protein function. The relative contributions of these mechanisms have been debated since King and Wilson proposed that functional species differences are largely the result of differential gene expression (King and Wilson 1975). Recent studies have confirmed the key role that altered gene expression plays in modifying body form or pattern (Shapiro et al. 2004; Löhr and Pick 2005; Prud'homme et al. 2006; Carroll 2008). However, structural changes in proteins also contribute to phenotypic adaptation (Jessen et al. 1991; Yokoyama et al. 1999; Hoekstra et al. 2006; Hoekstra and Coyne 2007). Recently, sweeping claims regarding the importance of each mechanism have been made by proponents on both sides of the debate (Hoekstra and Coyne 2007; Wray 2007), whereas others have argued that this dichotomy is arbitrary (Oakley 2007; Stern and Orgogozo 2008). In spite of this debate, few studies have examined the relative role that both mechanisms can play in shaping a single phenotype.

The visual system is ideal for investigating the molecular mechanisms of adaptation, because there is a direct link between genotype and phenotype (Bowmaker 1995; Yokoyama and Yokoyama 1996). Within the retina, spectral sensitivity is determined by visual pigments, which are composed of an opsin protein bound to a light-sensitive chromophore (Wald 1968). This opsin–chromophore interaction determines the peak spectral sensitivity of each visual pigment. Numerous studies have demonstrated that visual pigment sensitivities are tuned to the local light environment by amino acid substitutions in opsin proteins (Bowmaker 1995; Yokoyama and Yokoyama 1996; Yokoyama et al. 1999; Hunt et al. 2001; Terai et al. 2002; Carleton et al. 2005a; Sugawara et al. 2005; Terai et al. 2006; Seehausen et al. 2008). Consequently, sensory adaptation via changes in opsin gene coding sequence has become a classic example of molecular adaptation.

However, fish have numerous opsin genes that have arisen through tandem gene duplications. These duplicate opsin genes have diverged to produce visual pigments that absorb maximally across the full spectral range, from the ultraviolet to the red (reviewed in Hofmann and Carleton (2009)). Recent work in cichlids and other taxa has demonstrated that differential expression of these opsin genes may generate large changes in visual sensitivity (Carleton and Kocher 2001; Spady et al. 2006; Carleton et al. 2008; Shand et al. 2008). Typically, these studies have examined populations of one species, or of closely related species, but have not evaluated the relative importance, and adaptive significance, of spectral tuning via differential gene expression across many divergent species.

The haplochromine cichlids of the East African rift lakes are well suited for

addressing this question. They are a classic example of adaptive radiation and rapid speciation (Fryer and Iles 1972; Greenwood 1974; Schluter 2000; Streelman and Danley 2003; Kocher 2004; Seehausen 2006). Hundreds of new species have evolved in Lake Malawi within the past 1–2 million years and within a mere 15,000–120,000 years in Lake Victoria (Meyer et al. 1990; Genner et al. 2007). These two haplochromine radiations provide a large number of closely related, yet ecologically and morphologically divergent, species. Furthermore, these two lakes differ dramatically in their light environment (Carleton et al. 2006). Lake Malawi is one of the deepest and clearest freshwater lakes in the world, with clarity similar to that of marine environments (Muntz 1976). In contrast, Lake Victoria is relatively turbid, with long wavelength–shifted transmission and considerable variation in both clarity and transmission among geographic localities (Seehausen et al. 1997). Studies have demonstrated repeatedly that selection is acting on the visual systems of cichlids in both lakes (Sugawara et al. 2002; Terai et al. 2002; Carleton et al. 2005a; Spady et al. 2005; Sugawara et al. 2005; Terai et al. 2006; Seehausen et al. 2008).

In this study, we use these two replicate cichlid radiations to (*i*) examine how changes in opsin gene expression contribute to the remarkable diversification of cichlid visual systems, (*ii*) test whether changes in opsin gene expression are adaptive, and (*iii*) compare the relative roles that differential opsin gene expression and changes in protein coding sequence play in the diversification of cichlid visual systems.

Materials and Methods

Sampling and Ethics Statement

We collected 133 fish representing 52 cichlid species from Lake Malawi and 11 species/populations from Lake Victoria (Supplementary Table S2-1 [Appendix 1]). Lake Malawi species were wild-caught from the southern portion of Lake Malawi in 2005, either from the south side of Thumbi West Island or off Otter Point. Lake Victoria species were lab bred from wild-caught stocks and reared in a common garden laboratory environment at the Centre of Ecology, Evolution & Biogeochemistry of the ETH Institute for Aquatic Research in Kastanienbaum, Switzerland. Tanks were illuminated using daylight fluorescent light with a 12:12 light:dark cycle. Water temperature was kept constant at 24–26° C. All fish were raised on a mix of commercial flake food, given daily, and a blend of shrimp, peas, and Spirulina powder fed twice a week. Experimental tanks were part of a large recirculation system. All fish were sampled upon sexual maturity, then euthanized according to University of Maryland Institutional Animal Care and Use Committee (IACUC)-approved protocols (R-09-73).

Opsin Gene Expression

After euthanizing each fish, we dissected both retinas from their eyecups and immediately stored them in RNAlater (Ambion) until the time of analysis. Retinas were collected from adult fish greater than 6 months of age, when any ontogenetic changes would be complete (Carleton et al. 2008). These were collected during the late morning through the afternoon. Although cichlid opsin gene expression does

show diurnal variation, expression of cone opsin genes varies slowly and in synchrony (Halstenberg et al. 2005). Therefore, sampling time is not likely to impact the relative gene expression ratios we determined here.

The real-time qPCR methods we use to measure cichlid opsin expression follow those previously optimized for cichlid opsins (Carleton and Kocher 2001; Spady et al. 2006). In brief, RNA was extracted using commercially available kits (RNeasy, Qiagen) and reverse transcribed (Superscript III, Invitrogen). Real-time PCR reactions were run using opsin-specific TaqMan primers and probes that spanned the exon–exon boundaries. The recently diverged *RH2A*α and *RH2A*β opsin genes are genetically similar and produce visual pigments that differ in absorbance by only 10 nm (Spady et al. 2006). As in previous studies, we quantified them together (Carleton et al. 2005a; Spady et al. 2006; Carleton et al. 2008). Reactions for all six opsin classes were run in parallel. An internal standard containing a tandem array of segments from each opsin gene was used to calculate the reaction efficiency within each run. The relative expression of each opsin as a fraction of total cone opsin expression was then calculated from the reaction efficiency and critical cycle number (Carleton and Kocher 2001; Spady et al. 2006). Each reaction was run twice, and averages of both runs from all individuals of a species are reported.

We clustered species with quantitatively similar opsin gene expression profiles via hierarchical clustering. However, because multivariate methods such as hierarchical clustering are sensitive to factors with relatively larger values (Quinn and Keough 2002), we standardized the expression values of opsins expressed within single and double cones separately. To do this, we divided the relative expression of

each opsin by the combined expression of all other opsins within the same cone type (*SWS1*, *SWS2B*, and *SWS2A* for single cones; *RH2B*, *RH2A*, and *LWS* for double cones; see below for a justification of these assignments). This normalization procedure provides equal weighting to opsins expressed within single cones versus those expressed within double cones. We then used the normalized opsin expression data to calculate Euclidean distances between species and clustered them using Ward's method. We identified the optimal number of clusters resulting from this analysis using the Connectivity, Dunn, and Silhouette cluster validation indexes (Brock et al. 2008). Given a range of potential clusters, these indexes provide relative measures of support for each cluster size. Here, we tested for the presence of two to ten clusters. We implemented both hierarchical clustering and cluster validation statistics in the R package clValid (Brock et al. 2008).

Calculating Single- and Double-Cone Sensitivity

We calculated the average single- and double-cone sensitivities of all taxa in order to better understand how changes in gene expression might influence overall retinal sensitivity. First, we assigned opsin genes to cone types. Based on MSP data from 19 Malawian cichlid species (Carleton et al. 2000; Parry et al. 2005; Jordan et al. 2006), nine Victorian cichlid species (van der Meer and Bowmaker 1995; Carleton et al. 2005a), one Tanganyikan cichlid (Fernald and Liebman 1980), and the riverine cichlid, *Oreochromis niloticus* (Carleton et al. 2008), we have found that all cichlid single cones have a wavelength of maximum absorbance (λ_{max}) that is less than 460 nm, and all cichlid double cones have a λ_{max} that is greater than 460 nm. Based on the λ_{max} of heterologously expressed opsins from *O. niloticus* (Spady et al. 2006) and

M. zebra (Parry et al. 2005), this means that the *SWS1*, *SWS2B*, and *SWS2A* opsin genes are expressed in single cones, whereas *RH2B*, *RH2A*, and *LWS* are expressed in double cones.

To calculate average single- or double-cone sensitivities, peak spectral sensitivities for each opsin were weighted by the fraction of their expression in each cone type using the following equations:

$$\lambda_{max,\,S} = \frac{fSWS1\lambda SWS1 + fSWS2B\lambda SWS2B + fSWS2A\lambda SWS2A}{fSWS1 + fSWS2B + fSWS2A}$$

and

$$\lambda_{max,\,D} = \frac{fRH2B\lambda RH2B + fRH2A\lambda RH2A + fLWS\lambda LWS}{fRH2B + fRH2A + fLWS}$$

where f_i is the relative expression of the i^{th} opsin and λ_i is the λ_{max} of the same opsin (Spady et al. 2006; Carleton et al. 2008). We used previously published λ_{max} values from heterologously expressed *O. niloticus* opsins (*SWS1* = 360 nm, *SWS2B* = 425 nm, *SWS2A* = 456 nm, *RH2B* = 472 nm, *RH2A*α + *RH2A*β = 523 nm [mean], and *LWS* = 560 nm) (Spady et al. 2006). *O. niloticus* (Nile Tilapia) is considered an outgroup to both the Lake Malawi and Lake Victoria cichlid radiations (Kocher et al. 1995). We clustered and validated the number of single- and double-cone clusters using the same method as for the clustering of opsin expression values (see above).

Finally, although opsin expression and visual pigment sensitivity are tightly correlated (Carleton et al. 2005a; Spady et al. 2006; Carleton et al. 2008), these estimates of single- and double-cone sensitivity are not meant to suggest how colors are perceived (e.g., dichromacy vs. trichromacy). Rather, estimating single- and double-cone sensitivity allowed us to plot the data in a two-dimensional space to infer

how changes in gene expression influence overall retinal sensitivity in a quantitative manner. These single- and double-cone sensitivities were estimated based on two assumptions: (*i*) the visual pigment λ_{max} for each gene is the same for all species; and (*ii*) the chromophore is A1 (11-*cis* retinal) for all species. We have not attempted to estimate individual λ_{max} values for each gene in each species for several reasons. First, we have not sequenced all the genes from all species. Second, we do not know the effects of all the sites, which vary across each of the opsins, and so would not be able to predict the exact λ_{max}. However, based on the range of λ_{max} values that have been estimated from MSP of 30 different cichlid species from Lakes Malawi and Victoria, the variation in λ_{max} is relatively small: *SWS1* = 371±8 nm, *SWS2B* = 418±5 nm, *SWS2A* = 455±5 nm, *RH2B* = 482±5 nm, *RH2A* = 528±6 nm, and *LWS* = 565±9 nm (see Table 2 in Carleton et al. (2009)). Although there is larger variation in the *SWS1* and *LWS* visual pigments, in agreement with our sequence diversity (see Results section below), this variation would have a negligible effect on the placement of species in their respective opsin expression clusters. Therefore, a reasonable approximation is to use the same λ_{max} for each gene in all species. Similarly, we have neglected any effects of chromophore switching from A1 to A2. Malawian cichlids utilize primarily A1 chromophore. However, Victorian cichlids do show some evidence of A2 usage. A complete chromophore switch causes small shifts for *SWS1* (15 nm), *SWS2B* (7 nm), and *SWS2A* (10 nm), but larger shifts for *RH2B* (19 nm), *RH2A* (35 nm), and *LWS* (60 nm) based pigments (Carleton et al. 2006). It is more typical for the chromophore to be an A1/A2 mixture, which would decrease the size of these shifts. The net effect of A2 expression would be to push the double-cone

estimates for Victorian cichlids to longer wavelengths. This would stretch the long-wavelength cluster, but would never cause Victorian species to shift into the shorter-wavelength clusters. Further studies are needed to quantify chromophore usage in wild-caught fish, as this could be important for actual visual sensitivities.

Ecological Correlations within Lake Malawi

We used the phylogenetic comparative method (Felsenstein 1985) to test the hypothesis that opsin gene expression and the resulting single- and double-cone sensitivities differ adaptively among Lake Malawi cichlids with different foraging modes or macrohabitat preferences. Because of the lack of a resolved species-level phylogeny for this group, we used three different phylogenetic hypotheses for our analyses, (*i*) a mitochondrial gene tree reconstructed from 1,247 bp of mtDNA, (*ii*) a generic tree illustrating the purported taxonomic relationships among the genera sampled, and (*iii*) a star tree in which the *mbuna* and non-*mbuna* clades were collapsed into polytomies, representing their rapid radiation from a common ancestor (Figures S1-2). Our mtDNA phylogeny utilized 1,247 bp of mtDNA corresponding to 934 bp of the protein-coding gene *ND2* and 313 bp of the mitochondrial control region. Previously published sequences were available for approximately half the taxa sampled; the remaining sequences were obtained via PCR and direct sequencing following standard protocols (Kocher et al. 1995; Lee et al. 1995) (Table S2-7 [Appendix 1]). Multiple sequence alignment for the final sequence set was performed using the L-INS-i strategy of the program MAFFT (Katoh et al. 2002) and then edited manually. We then used MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001) to generate a phylogeny of all Lake Malawi species sampled. We specified the

GTR + Γ + G model of nucleotide substitution for these analyses following likelihood ratio tests in the program MrModeltest v2.2 (Nylander et al. 2004). We treated all model parameters as unlinked with a flat prior probability distribution. We performed three independent runs of four Metropolis-Hastings Coupled Monte Carlo Markov chains (MCMCMC), three hot and one cold. Each chain was run for 1,000,000 generations and trees were sampled every 1,000 generations. Posterior probability values for the resulting 50% majority rule consensus tree were estimated after discarding the first 10% of trees as burn-in.

We tested hypothesis that opsin gene expression varies adaptively among Lake Malawi using phylogenetic ANOVA implemented in the program PDSIMUL v2.0 (Garland Jr. et al. 1993). Null distributions of F-statistics for ANOVA, corrected for phylogenetic nonindependence, were generated by simulation (n = 1,000) of relative opsin gene expression levels and estimated single- and double-cone λ_{max} values across the three trees listed above. These simulations followed an unbounded Brownian motion model of character evolution. All statistical analyses were performed using the stats and PHYLOGR (Díaz-Uriarte and Garland Jr. 2007) packages in the program R v2.6.2.

Spectral Measurements

We measured the transmission properties of waters from Lakes Malawi and Victoria in the field. In Lake Malawi, the water attenuation coefficient as a function of wavelength was determined at two locations: (*i*) Zimbawe Island, a rocky outcrop with a maximum depth of 40 m, and (*ii*) the southern side of Thumbi West Island, a sheltered bay with a maximum depth of 15 m. A set of ten irradiance measurements

were taken from a series of depths (0, 1, 3, 5, 7, 10, 15, and 20 m at Zimbawe and 0, 1, 3, 7, and 10 m at Thumbi West) using Subspec, a submersible Ocean Optics (USB 2000) spectrometer fitted with a 100-mm fiber and a cosine correcter. These data were used to determine the slope (k, attenuation coefficient) and intercept (b) of a plot of $\ln(I_d/I_0)$ versus depth (d), where I_0 is the initial, full-spectrum irradiance, and I_d is the irradiance at depth. Transmission (T) at 2 m depth was then calculated using the equation $T = e^{(k*d+b)}$. Relative irradiance was then calculated by multiplying T by I_0.

Victorian water measurements were taken three locations: (*i*) Makobe Island, a relatively clear location, (*ii*) Python Island, a turbid location, and (*iii*) Luanso Island, an extremely turbid location. Transmission was measured at a depth of 2 m for all three locations, using an AvaSpec 2048 212 spectrophotometer with a 10 m fiber cable (100 mm) and SpectraWin 4.16 software (Avantes). Measurements were taken in the shade, between 8h30 and 9h00 in the morning. Irradiance was then calculated by multiplying T by I_0. The same I_0 (from Zimbawe) was used for both Malawi and Victoria to remove any daily variation and focus only on differences in water properties.

Calculating Relative Quantum Catch

We estimated the quantum catch (Q) that a visual pigment containing each opsin gene would have at each location in Lake Malawi and Victoria using the following equation:

$$Q = \int I(\lambda) T_W(\lambda,d) R(\lambda) d\lambda$$

where $I(\lambda)$ is the incident solar irradiance at the surface (measured at Zimbawe Rock), $T_W(\lambda,d)$ is the light transmission of the water to a depth (d = 2 m), and $R(\lambda)$ is the

photoreceptor absorption calculated using equations from Govardovskii et al. (Govardovskii et al. 2000). Because we were interested in the relative quantum catch each opsin gene would produce, we normalized the quantum catch for each visual pigment by the sum of the quantum catches from all visual pigments (this also removed intensity differences across geographic regions). Unpublished data suggest that ocular media are not limiting (e.g., species that express the UV-sensitive [*SWS1*] opsin have UV-transmitting lenses). Therefore, the potential influence of ocular media was not included in this estimate.

Ecological Correlations within Lake Victoria

We tested the hypothesis that that relative quantum catch of *SWS2B*-based visual pigments vary adaptively among cichlids from Lake Victoria once again using the phylogenetic comparative method. We first used Secchi disk readings (Table S2-4 [Appendix 1]) to classify the Lake Victoria cichlid population based on whether they came from clear (>150 cm) or turbid locations (<150 cm). Because we did not have measurements of the light environment from all locations, we used the attenuation coefficient from Makobe to represent clear water and from Python to represent turbid water. The mean depth each taxon inhabits at the location where it was collected was used to calculate the transmission and relative irradiance. We then calculated the relative quantum catch that an *SWS2B*-based visual pigment would have in this light environment using the equation described above.

To test whether *SWS2B* expression was correlated with visual pigment quantum catch (Table S2-4 [Appendix 1]), we used Felsenstein's independent contrasts method (Felsenstein 1985) as implemented in the PDAP v1.08 (Midford et

al. 2003) module of Mesquite v1.11 (Maddison and Maddison 2001). Because of the rapid nature of the Victorian radiation (< 100,000 y), we used a generic phylogeny for this analysis. To account for the presence of polytomies in this tree, we subtracted five degrees of freedom when calculating p-values for this analysis (see Figure S2-2 [Appendix 1]).

Opsin Sequence Diversity

We sequenced all seven cone opsin genes plus the rod opsin from five Lake Victoria taxa using previously published methods (Table S2-2 [Appendix 1]). Genomic DNA was isolated from fin clips and amplified using opsin-specific PCR primers (Carleton et al. 2000; Carleton and Kocher 2001; Parry et al. 2005). PCR products were gel or column purified and sequenced using PCR and internal primers. For all sequencing, we obtained at least 2X coverage and >95% of each gene's coding sequence.

Additional opsin sequences from previously published Lake Malawi and Victoria taxa were downloaded from GenBank (Table S2-2 [Appendix 1]). Since the *RH2A*α and *RH2B* gene sequences were missing for many of these taxa, we sequenced these genes for 18 taxa as well as any other missing or incomplete genes from genomic or cDNA stocks whenever possible (Table S2-2 [Appendix 1]). Sequences were assembled and edited using Sequencher (v4.9, Genecodes Corp.). Consensus sequences were then aligned, and intronic regions were removed. Previously published alignments between each cichlid opsin and bovine rhodopsin were used to identify amino acid substitutions that fell in the putative transmembrane and retinal binding pocket regions (Carleton et al. 2005b). Substitutions were then

examined to determine whether they were between amino acids with different physical properties. These properties were nonpolar hydrophobic, polar uncharged, polar acidic, and polar basic. This approach was chosen because of previous work that suggests statistical tests of selection in opsins can be misleading (Yokoyama et al. 2008). To rule out the possibility that the changes we observed were due to differences in the mutation rates of different opsins, we used MEGA v4.0 (Tamura et al. 2007) to calculate average pairwise rate of synonymous substitution (D_S), and nucleotide diversity (pi, or π) for each opsin.

Results

Opsin Expression Profiles

We quantified opsin gene expression in 54 wild-caught taxa from Lake Malawi and 11 lab-reared taxa from Lake Victoria (Tables S1-1 and S1-2). Cichlids have one rod opsin gene (*RH1*) and six functionally and genetically distinct classes of cone opsin: *SWS1* (ultraviolet, or UV), *SWS2B* (violet), *SWS2A* (blue), *RH2B* (blue-green), *RH2A* (green), and *LWS* (red) (Parry et al. 2005; Spady et al. 2006; Carleton et al. 2008). (As in previous cichlid studies, we group expression of the functionally and genetically similar *RH2A*α and *RH2A*β together (Carleton et al. 2005a; Spady et al. 2006; Carleton et al. 2008)). Cichlid retinas are highly organized, and the shorter-wavelength-sensitive (SWS) opsins are expressed in morphologically distinct single-cones, whereas the longer-wavelength RH2 and *LWS* genes are expressed in double-cones. Cichlids from Lake Malawi had diverse expression profiles that collectively expressed all six cone opsin genes (Figure 2-1). These expression profiles formed

three distinct clusters (Figure 2-2A) with support based on multiple cluster validation statistics (Table S2-8 [Appendix 1]). Members of the *mbuna* clade predominantly expressed the shorter-wavelength classes of opsin genes: all species sampled expressed *SWS1* or *SWS2B* opsins in their single cones, and fewer than half of these species (12/26) expressed the longer-wavelength *LWS* opsin in their double cones. Non-*mbuna* collectively expressed all three SWS opsins in their single cones, although the overwhelming majority of the species sampled (23/26) expressed *LWS* in their double cones (Table S2-1 [Appendix 1]). In both lineages, we found examples of closely related species that expressed different subsets of opsin genes, suggesting that sister taxa could differ significantly in visual sensitivity (Figure S2-1 [Appendix 1]). Such differences occurred in 12 of the 14 genera in which we sampled multiple species, and included genera as diverse as *Tropheops*, *Melanochromis*, *Protomelas*, *Dimidiochromis*, and *Rhamphochromis*.

Cichlids inhabiting Lake Victoria collectively expressed four different opsin classes (Figure 2-1), and their expression profiles fell within a single cluster (Figure 2-2A). None of the taxa that we examined expressed more than trace amounts of *SWS1* or *RH2B*. All of the Victorian species expressed *SWS2A* in their single cones and *RH2A* and *LWS* in their double cones. Several taxa also expressed *SWS2B* in their single cones, and *SWS2B* expression was variable, even among conspecifics from different geographic localities (rocky islands). We therefore treated each localized population as a distinct group in subsequent analyses (Table S2-1 [Appendix 1]).

To examine how changes in gene expression might shape overall retinal

sensitivity, we used data from reconstituted cichlid visual pigments (Spady et al. 2006) to estimate average single- and double-cone sensitivities for each species (Carleton et al. 2008). The estimated single- and double-cone sensitivities of Malawian taxa fell into three distinct groups sensitive to short-, middle-, and long-wavelength regions of the visible light spectrum (Figure 2-2B). These groups correspond directly to the gene expression clusters (Figure 2-2A) and were also supported by multiple cluster validation statistics (Table S2-8 [Appendix 1]). Although there was some variation in single- and double-cone sensitivities within Lake Victoria, all Victorian taxa fell into the long-wavelength group.

Ecological Factors Driving Divergent Opsin Expression: Lake Malawi

To test whether changes in gene expression were adaptive, we compared mean opsin expression and estimated photoreceptor sensitivity among Lake Malawi cichlids with different foraging and habitat preferences; however, in order to account for the statistical correlation between species due to their evolutionary history, we first generated several phylogenetic hypotheses to include in this analysis. The results of our mitochondrial phylogenetic analysis are presented in a final 50% majority-rule consensus tree in Figure S2-2A [Appendix 1]. As in other phylogenies of this group, internal branch lengths are extremely short, indicating simultaneous or near-simultaneous divergence from a common ancestor (Moran and Kornfield 1993; Parker and Kornfield 1997; Albertson et al. 1999; Won et al. 2006). Although this tree exhibits numerous polytomies, the resolved nodes generally have high posterior probability support. We also used two additional phylogenetic hypotheses for this group, including one resolved at the genus-level and another resolved at the clade-

level (Figure S2-2 [Appendix 1]).

Using phylogenetically controlled comparative methods, we found that the *SWS1* opsin gene was differentially expressed among Lake Malawi cichlids with different foraging preferences (phylogenetic ANOVA, $F_{4,45} = 7.647$, $p = 0.007$, Table S2-3 [Appendix 1]). *SWS1* expression was highest among species foraging on zooplankton, phytoplankton, and algae, and lowest among species foraging on fish or benthic invertebrates ($F_{1,52} = 23.91$, $p = 0.003$, Figure 2-3A). Up-regulation of *SWS1* also resulted in estimated single-cone sensitivities that differed among these species (phylogenetic ANOVA, $F_{4,45} = 9.065$, $p = 0.002$). Cichlids foraging on plankton and algae typically exhibited single-cone sensitivities peaking between 360 and 400 nm, such that they would be more sensitive to ultraviolet (UV) light than either piscivores or benthivores. *SWS1* was the only opsin significantly associated with foraging preferences. We did not observe significant differences in opsin gene expression or single- and double-cone sensitivities among cichlids from different habitats (rock, sand, intermediate, pelagic, or weeds; see Table S2-3 [Appendix 1]).

Several authors have addressed the problem of accounting for uncertainty in the phylogenetic relationships among taxa in comparative analyses (Grafen 1989; Martins and Garland Jr. 1991; Purvis and Garland Jr. 1993; Losos 1994; Garland Jr. and Díaz-Uriarte 1999). Due to the rapid radiation of the Lake Malawi cichlid species flock, and the lack of a clear, species-level phylogeny for this group (Moran and Kornfield 1993; Parker and Kornfield 1997; Albertson et al. 1999; Won et al. 2006), all polytomies were generally assumed to represent true simultaneous or near-simultaneous speciation events (e.g., "hard" polytomies). When the assumption of

hard polytomies holds, there should be no inflation in Type I error rates (Martins and Garland Jr. 1991). However, in order to account for the possibility of inflated Type I error rates due to over estimation of the true number of degrees of freedom from unresolved nodes, we specified short (0.25) branch lengths for all taxa emanating from polytomies and long (1.0) branch lengths for all taxa emanating from a bifurcation in each of our analyses. This correction effectively weights the results of contrasts between taxa separated by a bifurcation while reducing the influence of contrasts between taxa from a shared polytomy (Loh et al. 2008). The end result is a more conservative analysis akin to a reduction in the degrees of freedom for each polytomy when calculating significance values.

Phylogenetic studies that used very large numbers of nuclear gene markers have been able to resolve the Lake Malawi phylogeny with high statistical support at the genus level (Albertson et al. 1999; Allender et al. 2003; Kidd et al. 2006). These studies consistently found support for morphologically defined genera; however, the true phylogenetic relationships between genera remain unknown, and no large-scale study of both mbuna and non-mbuna genera are currently available. To take this into account, we calculated nested ANOVAs, nesting species in genus and genus in clade, using only Lake Malawi genera of which we had sampled at least two species. This approach is highly conservative because it makes no assumptions about phylogenetic relationships above the genus level, and assumes that each feeding style has evolved just once within each genus. Using this highly conservative approach, we found strong trends for associations between single cone sensitivities and feeding mode,

consistent with our previous analysis (nested ANOVA, $F_{19,12} = 2.339$, $p = 0.068$). Nested ANOVA was implemented in SPSS v17.0.

Ecological Factors Driving Divergent Opsin Expression: Lake Victoria

Although we sampled Lake Victoria taxa with a similar diversity of foraging preferences (e.g., planktivores, algivores, benthic foragers, and piscivores; see Table S2-1 [Appendix 1]), there was a complete absence of *SWS1* opsin expression among these cichlids, and all taxa fell into a single expression cluster (Figure 2-3B). These findings suggest that foraging preferences are not likely to be a major driver of opsin expression in the Victorian species that we sampled. However, photic environment is known to influence visual sensitivities among populations and species of cichlids from this lake (Carleton et al. 2005a; Terai et al. 2006; Seehausen et al. 2008). Therefore, we examined whether variation in the light environment between sampling sites could explain the pattern of gene expression that we observed.

We measured light transmission at three representative localities in Lake Victoria. We found that there was considerable variation between localities, with transmission decreasing and shifting to longer (redder) wavelengths from the open water site of Makobe to the sites of Python and Luanso, which were increasingly farther up the inlet of the Mwanza Gulf (Figure 2-4A). We then calculated how much of the available light a visual pigment composed of each opsin protein would capture at these different locations. In these spectrally narrow waters, quantum catches varied by almost four orders of magnitude (Figure 2-4B). *SWS2A*- and *LWS*-based visual pigments were predicted to have the greatest quantum catch in the single and double cones, respectively, whereas *SWS1*-based visual pigments would have virtually no

quantum catch (Figure 2-4B). *SWS2B*-based visual pigments would capture some of the available light in the relatively clear waters of Makobe, but very little at the other two, more turbid locations.

Finally, we used water clarity and population-specific depth preferences to predict the quantum catch that an *SWS2B*-based visual pigment would have at the site where each taxon was originally sampled (Tables S1-1 and S1-4). We found that *SWS2B* opsin gene expression was positively correlated with predicted quantum catch (Figure 2-3B, Felsenstein's independent contrasts: $r^2 = 0.456$, $F_{1,4} = 7.543$, $p = 0.023$), suggesting that *SWS2B* expression is increased in environments where it is predicted to capture more of the available light.

In the spectrally broad and relatively homogenous environment of Lake Malawi (Figure 2-5A), the estimated quantum catches do not vary appreciably between the two locations that we sampled (Zimbawe Rock, a deep, open-water site, and Thumbi West Island, a sheltered bay). Further, quantum catches vary by less than a single order of magnitude across opsin classes (Figure 2-5B). This finding suggests that environmental light is not likely to be a major driver of opsin gene expression in the species that were sampled from Lake Malawi.

Changes in Opsin Coding Sequences

Several previous studies have documented the action of selection on different cichlid opsin genes (Sugawara et al. 2002; Terai et al. 2002; Carleton et al. 2005a; Spady et al. 2005; Terai et al. 2006; Seehausen et al. 2008). To complement those studies, we compared coding sequence diversity across the cone and rod opsins of ten species from Lake Victoria and 16 species from Lake Malawi (Table S2-5 [Appendix

1]). We focused on substitutions between amino acids with different chemical properties in the transmembrane and retinal binding pocket regions of the protein because changes in these regions are most likely to alter visual pigment sensitivity. We found that the number and nature of amino acid substitutions varied considerably across opsin classes (Figure 2-6B). Among species sampled from Lake Malawi (Figure 2-6C), the greatest diversity of functionally critical sites was found in the *SWS1* (UV-sensitive) opsin, which had seven variable transmembrane sites, of which three were in the retinal binding pocket. Both the *LWS* and *RH1* opsins exhibited four variable transmembrane sites, of which three and two, respectively, were in the retinal binding pocket. Among cichlids from Lake Victoria (Figure 2-6D), the number of functionally important sites was highest for the *LWS* opsin, which had five variable transmembrane sites, of which three were in the retinal binding pocket. Several of these substitutions were at sites that have been previously demonstrated to shift the spectral sensitivities of visual pigments. For example, microspectrophotometry suggests that the Lake Malawi cichlids *Metriaclima zebra* and *Psuedotropheus acei* have *SWS1* visual pigments that differ in their wavelength of maximum sensitivity (λ_{max}) by ~ 10 nm, confirming the variable spectral sensitivity of the *SWS1* pigments among Lake Malawi cichlids (see also Smith et al. (2010). This variation in spectral sensitivity correlates with amino acid substitutions within the retinal binding pocket of the *SWS1* opsin that differ in polarity. *M. zebra*, whose *SWS1* opsin has a λ_{max} of 368 nm (Carleton et al. 2008), has the combination of a serine at site 114, a threonine at site 160, and a threonine at site 204 (S114/T160/T204). *P. acei*, whose *SWS1* opsin has a λ_{max} of 378 nm (Parry et al. 2005), has an alanine at site 114, an alanine at site

160, and an isoleucine at site 204 (A114/A160/I204). In the *LWS* opsin, variation from an alanine to a serine at site 164 (A164S) is known to long-wavelength shift the λ_{max} of the *LWS* opsin (Asenjo et al. 1994), and variation in the *LWS* sequence of *Pundamilia pundamilia* has been shown to produce a visual pigment that is shorter wavelength that that of *P. nyerei* by 3 – 15 nm (Carleton et al. 2005a; Terai et al. 2006) (Table S2-6 [Appendix 1]). These finding suggest that the opsin sequence variation we observe generates functional variation in the corresponding visual pigments. Longer wavelength shifts occur in species which inhabit deeper waters where the light is relatively more red-shifted (Seehausen et al. 2008). The observed number of functional substitutions was independent of the number of synonymous changes and of overall nucleotide diversity (Figure S2-3 [Appendix 1]).

Discussion

We present a comprehensive analysis of opsin gene expression in over 60 different species of cichlids from Lakes Malawi and Victoria. We found that changes in opsin expression can generate diverse sets of visual systems. We also demonstrated that these changes in gene expression are adaptive and are shaped by foraging preferences and the local light environment. In addition, we examined coding sequence variation across the full complement of opsin genes. We found that diversity in functionally important regions is not distributed equally. Instead, diversity is highest in the opsin genes that code for the shortest- and longest-wavelength visual pigments. Although numerous studies have demonstrated the importance of changes in opsin coding sequence to visual adaptation in cichlids, only

one study addressed adaptive changes in opsin gene expression, and this was only for a limited number of closely related species (Carleton et al. 2005a). Our results suggest a model of sensory adaptation where evolutionary changes in both expression and coding sequence work in concert to shape visual pigment sensitivity.

Visual System Diversity

We found that cichlids inhabiting the spectrally broad light environment of Lake Malawi had remarkable visual diversity and collectively expressed all six cone opsin genes. Although opsin expression was labile and could differ among closely related species, some structure emerged when the two major lineages within Lake Malawi were compared. Members of the *mbuna* or rock-dwelling clade predominantly expressed the shorter-wavelength classes of opsin genes in both single and double cones. Non-*mbuna* (sand-dwelling or pelagic species) collectively expressed all six opsins, but the middle- and longer-wavelength classes were predominant. Cichlids inhabiting the turbid waters of Lake Victoria express only four different classes of cone opsin. The shortest-wavelength single- and double-cone opsin genes were never expressed, and the longest-wavelength genes were expressed ubiquitously.

When we estimated single- and double-cone sensitivities based on patterns of opsin expression, we found that the species fell into three distinct short-, middle-, and long-wavelength clusters. These clusters correspond well with the three ‘‘visual palettes’’ documented previously in these and other cichlid species using microspectrophotometry (MSP) (Fernald and Liebman 1980; Parry et al. 2005; Jordan et al. 2006; Carleton et al. 2008). Cichlids from Lake Malawi utilized every visual

palette, whereas all Victorian cichlids grouped with the Malawian long-wavelength one. Thus, our results suggest that regulatory changes in opsin gene expression have generated diverse sets of single- and double-cone sensitivities. This extent of visual diversity among so many closely related species is extraordinary.

Divergence in Opsin Expression is Adaptive

We found evidence that changes in gene expression contributed to sensory adaptation, both to enhance foraging and to adapt to differences in the photic environment. The *SWS1* opsin gene, which encodes a UV-sensitive visual pigment, was differentially expressed between cichlids from different trophic groups in the clear waters of Lake Malawi. Species feeding on plankton or algae typically exhibited single-cone sensitivities peaking at shorter wavelengths than piscivores or benthic foragers. Studies of several teleost species, including two of the cichlids examined in this study, have demonstrated that UV sensitivity can increase the efficiency of foraging on zooplankton and other small organisms (Loew et al. 1993; Novales-Flamarique and Hawryshyn 1994; Jordan et al. 2004). Additionally, many cichlids are opportunistic feeders, and several species have been observed to switch from foraging on algae to foraging on zooplankton or phytoplankton (McKaye and Marsh 1983). We found that expression of the *SWS1* opsin is highest precisely among cichlids foraging on these food sources (Figure 2-3A). Given that our comparative results are also supported by experimental and observational data, we believe that the observed differences in *SWS1* opsin expression are adaptive and that foraging may be a key driver of visual pigment diversity in Lake Malawi (Coddington 1988; Martins 2000).

Ambient light appears to have a strong influence on opsin expression in the spectrally narrow, longer-wavelength waters of Lake Victoria. We found that all of the Victorian species that we sampled exhibited similar expression profiles, with some variation in the expression of *SWS2B*. The predominant opsin genes expressed among these taxa—*SWS2A* (blue) in single cones, and *RH2A* (green) and *LWS* (red) in double cones—were predicted to produce visual pigments with the greatest quantum catches in all three of our representative light environments. However, our predictions also suggested that an *SWS2B*-based visual pigment (violet) would capture some of the available light in clear locations, but much less in turbid ones. *SWS2B* opsin gene expression varied across taxa, and this variation was positively correlated with predicted quantum catch. Taken together, our findings suggest that ambient light is driving opsin gene expression in Lake Victoria.

One potential limitation of our study was that the Malawian samples were wild-caught, whereas the Victorian samples were lab-reared in a common garden environment. Although lab rearing and light manipulations have been demonstrated to alter levels of opsin expression, photoreceptor abundance, and photoreceptor length (Wagner and Kroger 2000; Fuller et al. 2004, 2005; Shand et al. 2008), several lines of evidence suggest there is a large genetic component to opsin expression in cichlids. First, all three opsin expression clusters are observed in species raised in a common lab environment. In fact, the three opsin palettes of Lake Malawi were originally identified in lab-reared fish (Carleton and Kocher 2001; Parry et al. 2005), and all seven opsin genes are turned on in ontogenetic sequence in tilapia raised under laboratory conditions (Spady et al. 2006; Carleton et al. 2008). Second, genetic

crosses between cichlid species with different visual palettes found a significant genetic component to opsin expression (Carleton et al. 2010). Finally, direct comparisons of gene expression from wild-caught and lab-reared F_1 fish from the same populations in Lake Malawi suggest that, expression of the shortest-wavelength *SWS1* and *SWS2B* opsins is maintained in the lab, although the exact levels of gene expression may change for some opsins in some species (Hofmann et al. 2010b). In sum, we feel that the lab rearing of Victorian samples is unlikely to influence our overall finding that differences in gene expression are adaptive.

Potential for Speciation

The rapid changes in opsin gene expression that we observed among these closely related cichlid species are unprecedented in vertebrates. Differential gene expression among these species produces large shifts in spectral sensitivities (up to 100 nm) that could modify a species' view of conspecifics or the natural scene, and so modify species behavior. In Lake Victoria, changes in the coding sequence of the *LWS* opsin result in smaller shifts (5– 15 nm) in visual pigment sensitivity that are linked to differences in depth, water clarity, and male color (Carleton et al. 2005a; Terai et al. 2006; Seehausen et al. 2008). As a result, the *LWS* opsin gene is under strong selection and was shown recently to play a role in speciation in cichlids from Lake Victoria (Seehausen et al. 2008). Since these fine- scale changes are linked to speciation, it is likely that the large differences in visual pigment sensitivity generated through differential opsin expression could also play such a role in cichlids from both lakes.

Increased Diversity in the Longest- and Shortest- Wavelength Opsins

Opsin genes provide a clear example of how gene duplication and divergence in coding sequence can generate functional diversity in an adaptive phenotype (Yokoyama 2002). We found strong evidence for functional coding differences among species, though these were not distributed equally across the opsins. The greatest number of functional coding differences were in the cone opsin genes that produce visual pigments at the ends of the cichlid visual range— the *SWS1* (UV) and *LWS* (red) opsins—as well as in the *RH1* (rod) opsin. Since the rod opsin is the only opsin expressed in cichlid rods, rods cannot use the mechanism of differential gene expression to tune visual pigment sensitivity. Likewise, differential gene expression cannot extend spectral sensitivity beyond the boundaries set by the opsin genes that encode the shortest- and longest-wavelength visual pigments (because there are no shorter- or longer-wavelength genes to turn on). Therefore, all three of these genes must utilize coding sequence changes to alter visual pigment sensitivity. This pattern of sequence diversity is consistent with previous evidence that selection is acting on these three opsin genes (Sugawara et al. 2002; Terai et al. 2002; Spady et al. 2005; Terai et al. 2006).

A Model of Sensory Diversification

In this study, we examined the different contributions that changes in gene expression and coding sequence make to the diversification of cichlid visual systems. Our results suggest a model in which both proximate mechanisms contribute to visual pigment diversity. This model contains three main features: (*i*) Differential gene expression can generate large shifts in visual pigment sensitivity (30–100 nm) across

the combined opsin spectral range; (*ii*) coding sequence substitutions fine-tune visual pigment sensitivity (5–15 nm) around each opsin's ancestral sensitivity; (*iii*) changes in coding sequence are more prevalent in the opsins operating at the short- and long-wavelength ends of the visual range, where differential gene expression can no longer extend visual pigment sensitivity. Therefore, although tuning in the middle portion of the visible-light spectrum is achieved by shifts in opsin gene expression, tuning at the ends of the visible light spectrum is achieved via opsin sequence evolution.

This model suggests that changes in gene expression and changes in protein coding sequence work in concert to generate phenotypic diversity. The extent to which our model can be applied to the visual systems of other teleosts, other sensory systems, or other genetic pathways remains to be seen. However, we predict that phenotypes influenced by multiple paralogous genes are likely to show similar patterns of expression and coding sequence evolution. We are currently examining the visual systems of Lake Tanganyika cichlids and damselfish. These two radiations are older than those in this study by one and two orders of magnitude, respectively, and will provide further tests for how coding sequence and gene expression interact in shaping visual phenotypes. Finally, we are performing genetic crosses to identify the specific loci that are responsible for the changes in gene expression that we observe. Understanding the timescales over which structural and regulatory changes act, and understanding the loci underlying regulatory changes, will provide further insights into when and how they work in concert to generate adaptive phenotypic change.

Figures

Figure 2-1. Opsin gene expression from all species surveyed. Triangle plots illustrate the relationships between opsins within the single and double cones of Lake Malawi and Victoria cichlids. Bar graphs illustrate expression of the corresponding opsins and emphasize the qualitative differences between lakes. No Lake Victoria taxa express more than trace amounts of SWS1 or RH2B (open bar), and all express high level of LWS (filled bar).

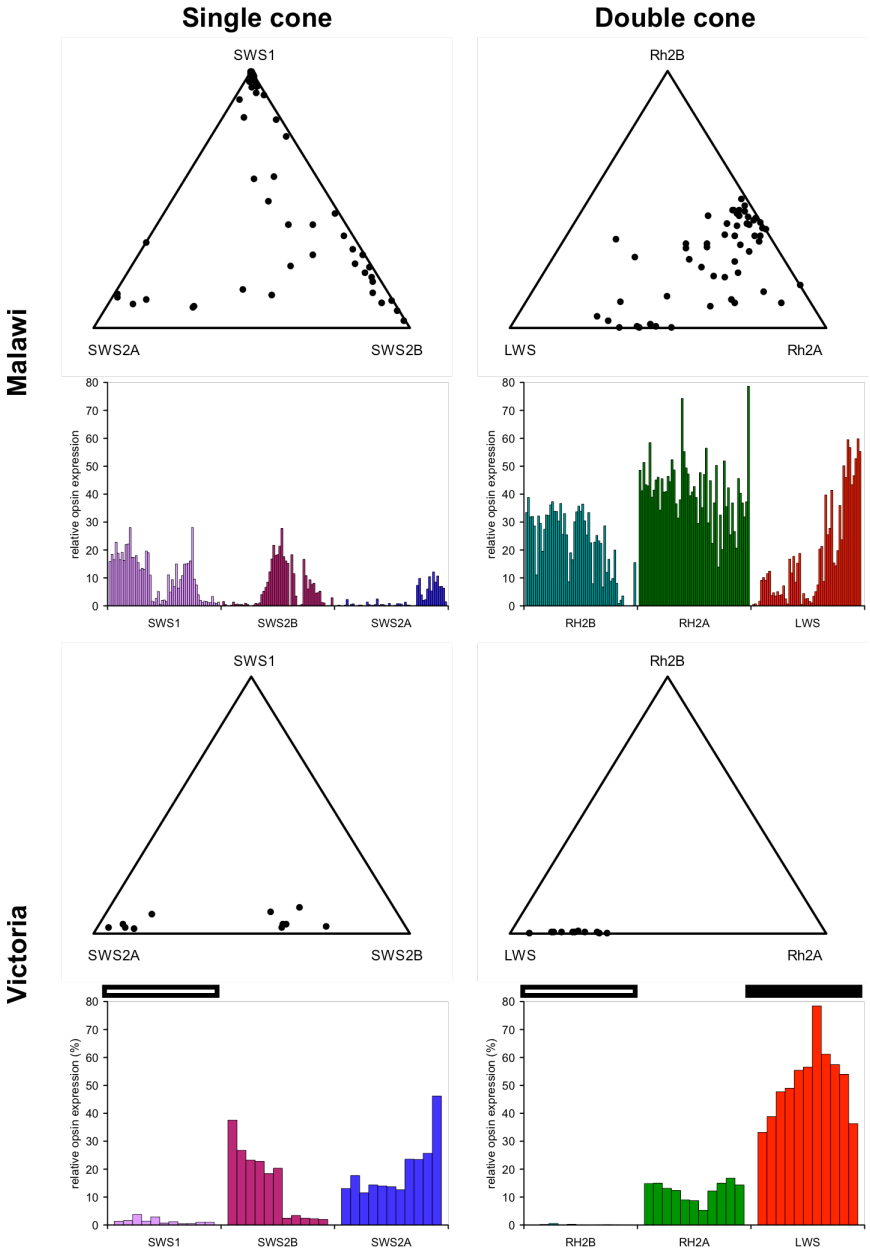


Figure 2-2. Gene expression profiles and single- and double-cone sensitivities form three clusters. (A) Hierarchical clustering of species' opsin expression profiles revealed three clusters: S, short wavelength; M, medium wavelength; L, long wavelength. (B) Estimates of Lake Malawi single- and double-cone sensitivities suggest that these three clusters correspond to visual palettes sensitive to short-, medium-, and long-wavelength portions of the visible light spectrum. Lake Malawi cichlids that are members of the mbuna clade are shown blue, non-mbuna are green, and Lake Victoria cichlids are red.

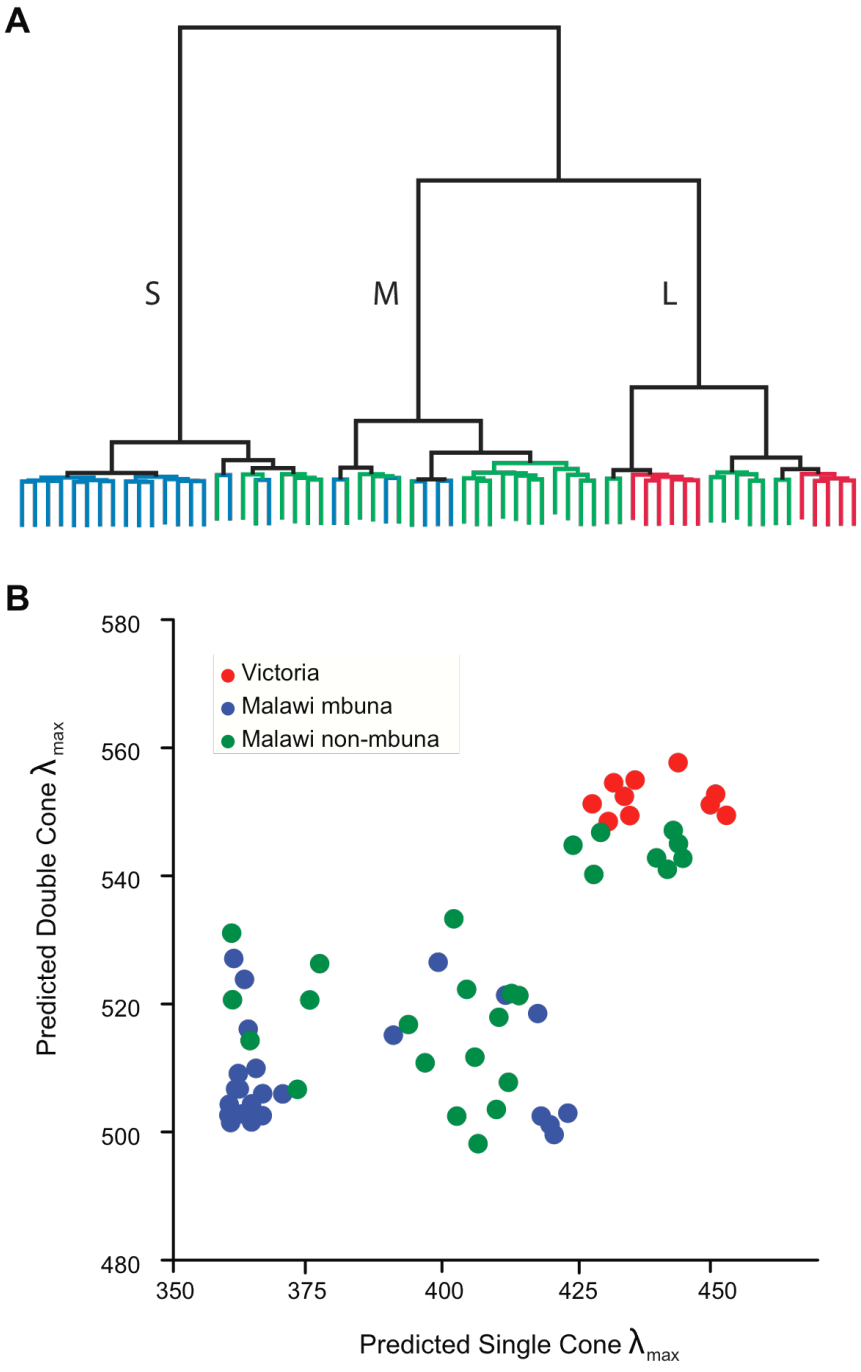


Figure 2-3. Selective pressures drive opsin expression with each lake. (A) Relative *SWS1* (ultraviolet) expression is higher among Lake Malawi cichlids that forage on plankton and algae. (B) Relative *SWS2B* expression is positively correlated with the predicted quantum catch of a *SWS2B*-based visual pigment from the clear and murky locations where Lake Victoria cichlids were sampled.

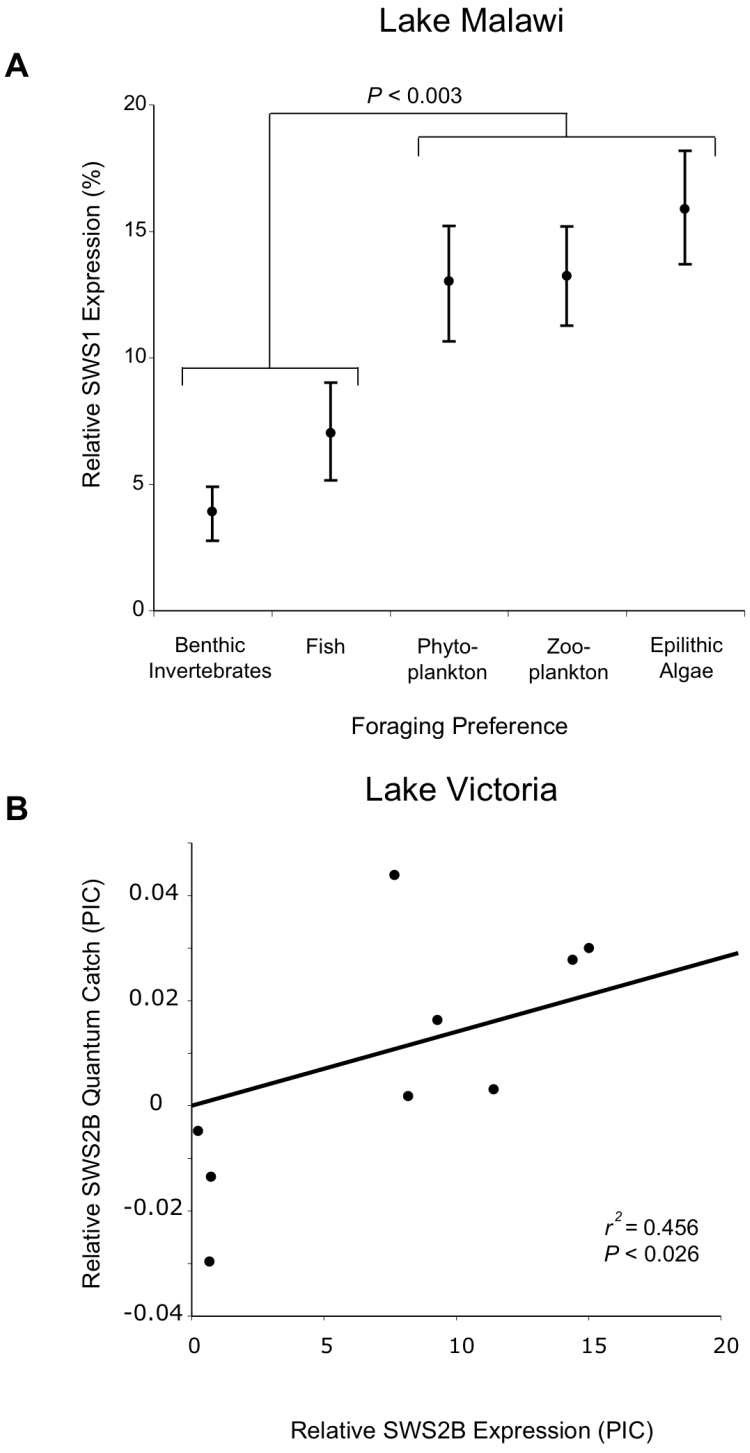


Figure 2-4. Visual pigment performance in Lake Victoria. (A) Relative irradiance at 2 m depth at three locations in Lake Victoria (Makbe Island, Python Island, and Luanso Island). (B) In Lake Victoria, estimated quantum catches vary over several orders of magnitude, both across visual pigments and geographic locations.

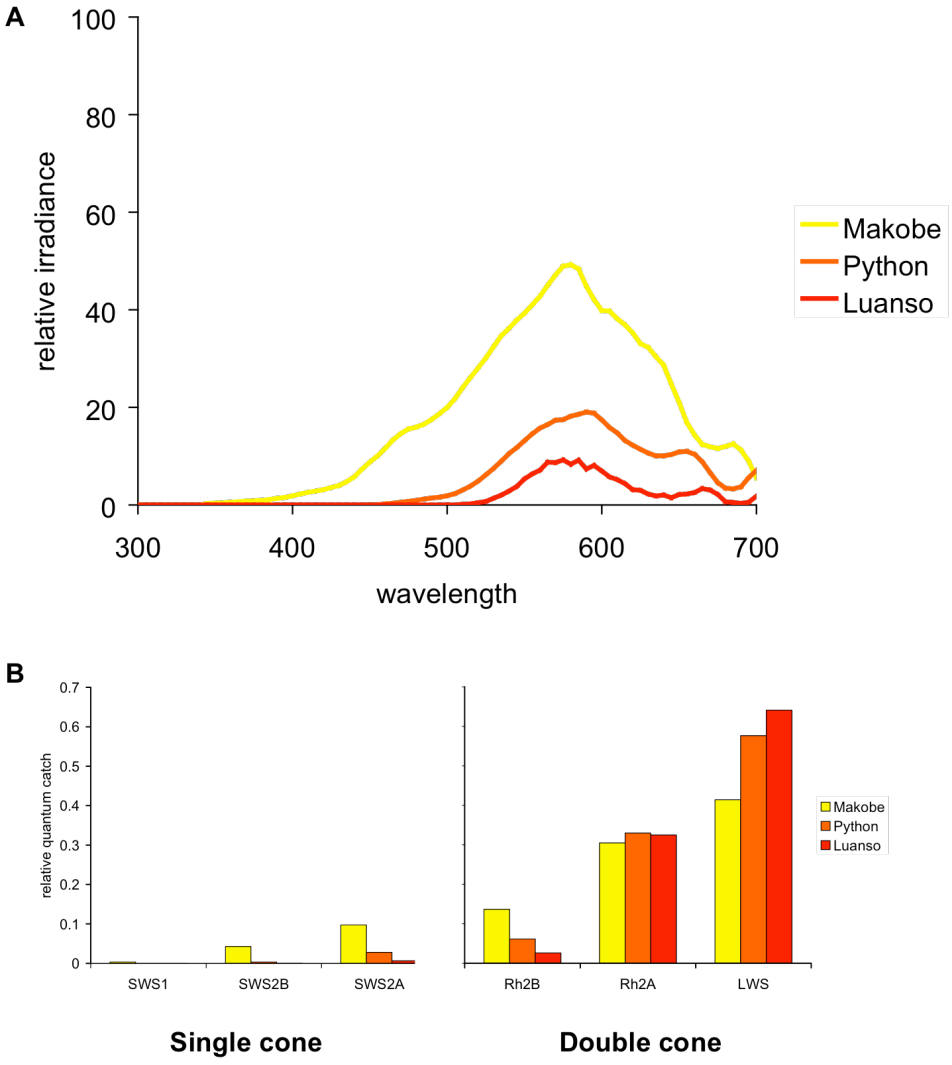


Figure 2-5. Visual pigment performance in Lake Malawi. (A) Relative irradiance at 2 m depth at two locations in Lake Malawi (Thumbi West and Zimbawe Rock). (B) In Lake Malawi, all visual pigments are estimated to have relatively similar, high quantum catches at both locations.

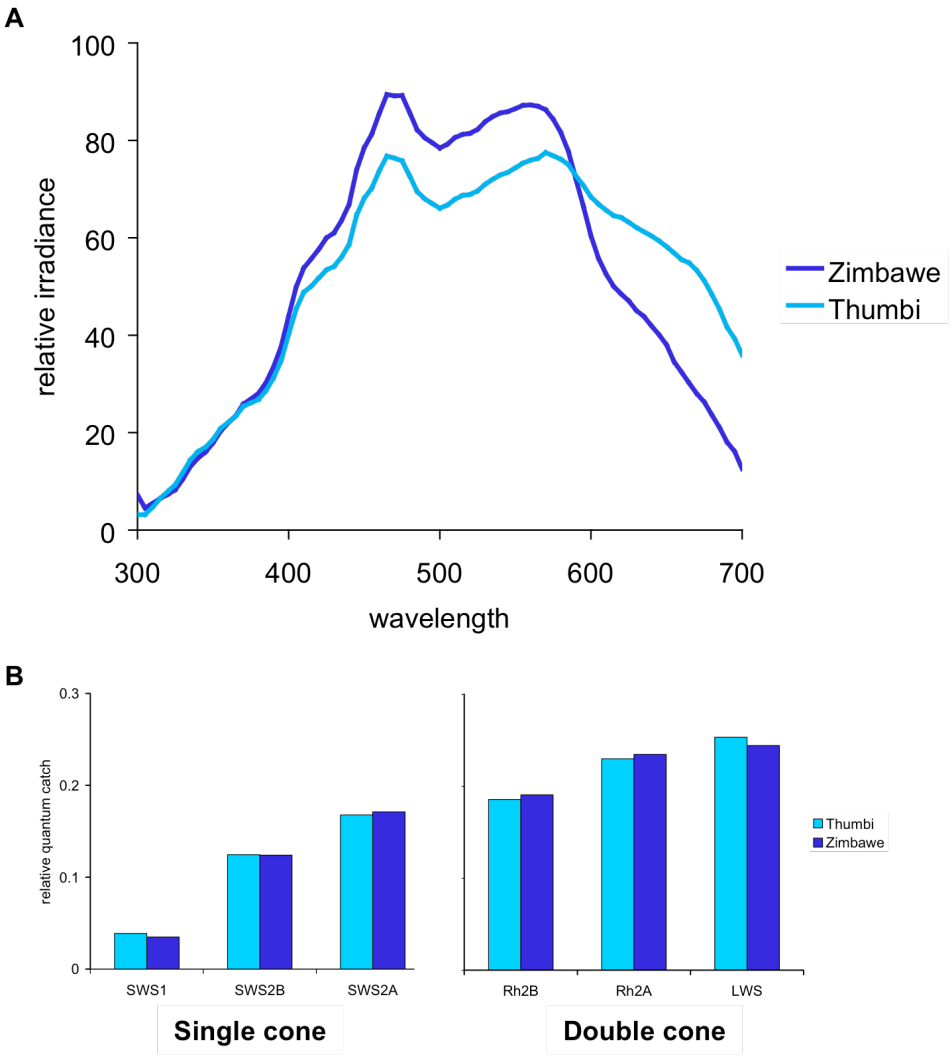
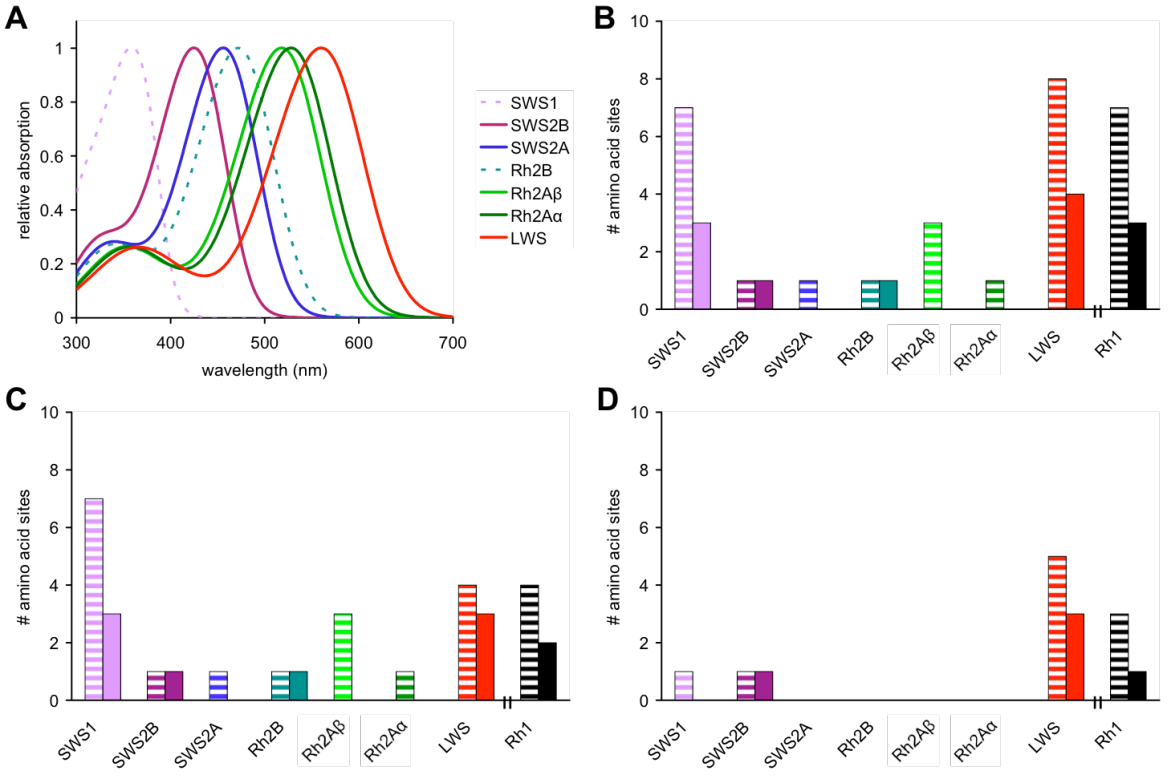


Figure 2-6. The shortest- and longest-wavelength opsins have the greatest sequence diversity. (A) Normalized absorbance values of all seven cichlid cone visual pigments, generated using the λ_{max} of each opsin in *Oreochormis niloticus*. Dotted lines represent opsins not expression in the Lake Vcitoria populations sampled. (B) Number of amino acid substitutions in the transmembrane regions (barred) and retinal binding pocket (solid) of each opsin class from all species surveyed. (C) Number of substitutions in Lake Malawi. (D) Number of substitutions in Lake Victoria. Only substitutions between residues with different chemical properties are shown.



Chapter 3:

Parallel Evolution of Opsin Gene Expression in African Cichlid Fishes

Kelly E. O'Quin, Christopher M. Hofmann, Hans A. Hofmann, Karen L. Carleton

See Appendix 2 for all supplementary tables (Tables 3-1 – S3-5) and figures (Figure S3-1) referenced in this chapter.

This chapter is published as:

O'Quin KE, Hofmann CM, Hofmann HA, Carleton KL (2010) Parallel Evolution of Opsin Gene Expression in African Cichlid Fishes. *Molecular Biology and Evolution* 27(12): 2839-2854.

Abstract

Phenotypic evolution may occur either through alterations to the structure of protein-coding genes or their expression. Evidence for which of these two mechanisms more commonly contribute to the evolution of a phenotype can be garnered from examples of parallel and convergent evolution. The visual system of East African cichlid fishes is an excellent system with which to address this question. Cichlid fishes from Lakes Malawi (LM) and Victoria together exhibit three diverse palettes of coexpressed opsins and several important protein-coding mutations that both shift spectral sensitivity. Here we assess both opsin expression and protein-coding diversity among cichlids from a third rift lake, Lake Tanganyika (LT). We found that Tanganyikan cichlids exhibit three palettes of coexpressed opsins that largely overlap the short-, middle-, and long-wavelength–sensitive palettes of LM cichlids. Bayesian phenotypic clustering and ancestral state reconstructions both support the parallel evolution of the short- and middle-wavelength palettes among cichlids from LT and LM. In each case, these transitions occurred from different ancestors that expressed the same long-wavelength palette. We also identified similar but distinct patterns of correlated evolution between opsin expression, diet, and lens transmittance among cichlids from LT and LM as well. In contrast to regulatory changes, we identified few functional or potentially functional mutations in the protein-coding sequences of three variable opsins, with the possible exception of the *SWS1* (ultraviolet) opsin. These results underscore the important contribution that gene regulation can make to rapid phenotypic evolution and adaptation.

Introduction

Phenotypic evolution may occur either through alterations to the structure of protein-coding genes or their expression. Mutations that alter the structure of protein-coding genes have long been known to underlie adaptive phenotypic differences between populations and species (Jessen et al. 1991; Hoekstra et al. 2006; Protas et al. 2006). However, recent work has provided abundant new evidence that mutations that alter the regulation or expression of genes also contribute to adaptive phenotypic evolution (Wittkopp et al. 2003; Shapiro et al. 2004). Evidence for which of these two mechanisms more commonly contribute to the evolution of a phenotype can be garnered from examples of repeated evolution either through parallelism or convergence (Gompel and Prud'homme 2009). For example, the parallel loss of pelvic spines among adaptively radiating sticklebacks has been achieved through recurrent mutations in the *cis*-regulatory region of *Pitx1* (Chan et al. 2010). This observation suggests that the evolution of pelvic spine loss in sticklebacks is biased toward regulatory mutations. Similar examples for protein-coding mutations also exist. For example, reduced pigmentation phenotypes have evolved repeatedly among vertebrates. In many cases, these convergent phenotypes arose through independent mutations within the protein-coding region of *Mc1r* (Mundy 2005; Gompel and Prud'homme 2009).

The visual system of African cichlids is an excellent model with which to study the roles of protein-coding and regulatory mutations during phenotypic evolution. Both protein-coding mutations and regulatory changes contribute to spectral sensitivity in these fishes (Hofmann and Carleton 2009; Hofmann et al. 2009)

[see Chapter 1]. Spectral sensitivity—or sensitivity to different wavelengths of light—is determined by the coding sequence and expression of several duplicated opsin genes. These opsins are expressed within distinct photoreceptor cells in the retina and, when combined with a light-sensitive chromophore, confer sensitivity to light (Wald 1935). Cichlids have 8 opsin genes, 7 used for bright light, or photopic, vision, and one used for dim-light, or scotopic, vision (Carleton 2009). These opsins are *SWS1* (ultraviolet [UV]), *SWS2B* (violet), *SWS2A* (blue), *RH2B* (blue–green), *RH2A*α and *RH2A*β (green), *LWS* (red), and *RH1* (dim light) (Spady et al. 2006). Among cichlid fishes from Lake Malawi (LM), closely related species can differ in their maximal short- and long-wavelength spectral sensitivity by as much as 100 nm (Jordan et al. 2006; Carleton 2009; Hofmann et al. 2009). These differences are highly correlated with discrete changes in opsin gene expression (Carleton and Kocher 2001; Hofmann et al. 2009). LM cichlids collectively coexpress three distinct opsin gene palettes, which generate visual pigment sets broadly sensitive to short- (*SWS1-RH2B-RH2A*), middle- (*SWS2B-RH2B-RH2A*), and long (*SWS2A-RH2A-LWS*)-wavelength spectra. The differential expression of these palettes is in part an adaptive response to divergent foraging preferences or diet (Hofmann et al. 2009). In contrast, cichlids from Lake Victoria (LV) collectively express only a single-opsin palette, the long-wavelength set (*SWS2A-RH2A-LWS*). However, these species do vary slightly in the expression of the *SWS2B* and *LWS* opsins. These smaller, continuous changes are an adaptive response to local differences in the light environment (Carleton et al. 2005a; Hofmann et al. 2009).

Opsin protein–coding mutations are also associated with the adaptive

evolution of spectral sensitivity in cichlids and other vertebrates (Yokoyama and Yokoyama 1996; Spady et al. 2005). In some cases, these protein- coding mutations are even associated with population divergence and speciation. For example, among cichlids from LV, polymorphisms in the protein-coding sequence of the *LWS* opsin are adaptively associated with local variation in the light environment, male color, and speciation (Terai et al. 2006; Seehausen et al. 2008). Among cichlids from both LM and LV, the majority of functional or potentially functional opsin sequence polymorphisms are found in the two opsins sensitive to the ends of the visible light spectrum (*SWS1* and *LWS*). This pattern is due to the inability of changes in gene expression to further tune spectral sensitivity outside of the spectral range of these two opsins as defined by their coding sequences. In contrast, shifts in gene expression predominately tune sensitivity across the middle portion of the visible light spectrum, where opsins of longer or shorter spectral sensitivity can be replaced with one another. Therefore, coding mutations are the only way to further shift spectral sensitivity at the ends of the visible light spectrum (Hofmann et al. 2009). Despite this observation, virtually all the cichlid opsins exhibit molecular signatures of natural selection (Sugawara et al. 2002; Spady et al. 2005), including those sensitive to the middle portion of the visible light spectrum; however, it is possible that in some cases, these estimates are too liberal (Yokoyama et al. 2008). Additionally, these polymorphisms are correlated with much smaller differences in spectral sensitivity, typically on the order of 5–15 nm (Carleton 2009; Hofmann et al. 2009).

Thus, African cichlids provide a unique system with which to investigate the

relative contribution that opsin regulatory and protein-coding mutations make to phenotypic evolution. However, cichlids from LM and LV form reciprocally monophyletic groups that are composed entirely of species from a single lineage, the Haplochromini (Salzburger et al. 2002; Koblmüller et al. 2008); see Figure 3-1). Cichlids from these lakes share very few opsin protein-coding polymorphisms in common, but at least some species share similar opsin expression profiles (*SWS2A-RH2A-LWS*) (Hofmann et al. 2009). But, due to the sister relationship of these two groups, it is unclear if this similarity is due to repeated evolution or shared ancestry. Therefore, here we assess opsin gene expression in 28 cichlids from a third nearby lake, Lake Tanganyika (LT). Cichlids from LT are both phylogenetically and phenotypically more diverse than cichlids from either LM or LV (Huber et al. 1997; Salzburger et al. 2002; Pollen et al. 2007). LT contains cichlids from many diverse lineages and tribes, including many older taxa that are ancestral to the LM and LV cichlid species flocks (Sturmbauer 1998; Takahashi 2003) (Figure 3-1); thus, cichlids from LT should provide a tractable system for identifying potential examples of repeated evolution in opsin gene expression. However, little is known of the spectral sensitivity of cichlids in LT. Although visual acuity has been documented for four species from the tribe Ecotodini (Dobberfuhl et al. 2005), actual retinal sensitivities have been measured for only a single LT cichlid, *Astatotilapia burtoni* (Fernald and Liebman 1980). Hence, among cichlids from LT, it is unclear whether retinal sensitivities are evolving primarily through opsin protein-coding mutations or regulatory changes. We hypothesize that similar opsin expression palettes will be present among cichlids from LM and LT because both these lakes have clear,

spectrally broad waters (Carleton et al. 2006) and both contain cichlids with parallel morphological and ecological adaptations (Kocher et al. 1993; Kassam et al. 2003).

In addition to repeated phenotypic evolution, we also assess phenotypic correlations between opsin expression and two factors associated with opsin expression divergence in cichlids, diet (Hofmann et al. 2009) and lens transmittance (Hofmann et al. 2010a). Among cichlids from LM, the *SWS1* (UV) opsin is up-regulated among species that forage on zooplankton and other microorganisms (Hofmann et al. 2009). This adaptation increases the ability of cichlids and other teleosts to find and capture zooplankton (Novales-Flamarique and Hawryshyn 1994; Jordan et al. 2004). Also among cichlids in LM, lens transmittance is positively correlated with both relative *SWS1* expression and the estimated sensitivity (λ_{max}) of single cone photoreceptors (Hofmann et al. 2010a). This correlation reveals that cichlids do not express opsins sensitive to wavelengths of light that their lenses ultimately filter before reaching the retina.

In summary, our goals were to: (*i*) test the hypothesis that similar opsin gene expression palettes have evolved repeatedly among African cichlids in LT and LM, and (*ii*) test for the presence of similar phenotypic correlations among opsin expression, diet, and lens transmittance. The repeated evolution of these opsin palettes would suggest that regulatory mutations have played an important role in the evolution of spectral sensitivity among African cichlids. Additionally, the independent evolution of one or more phenotypic correlations would implicate natural selection as one driver of opsin expression evolution in these fishes (Schluter 2000).

Materials and Methods

Sampling Tanganyikan Cichlids

We sampled 85 individual fish representing 28 different LT cichlid species. Half of these samples were collected as adult fish from LT near Kigoma, Tanzania, in 2004. The remaining species were purchased as wild-caught adult fish from a commercial supplier. Additionally, we also sampled adult fish from a laboratory strain of *Astatotilapia burtoni*. All eyes were collected at midday from full spectrum light-adapted animals. We noted the primary diet of each species following a survey of relevant literature sources (Taborsky et al. 1986; Yamaoka et al. 1986; Brichard 1989; Salzburger et al. 2002; Takahashi 2003; Duftner et al. 2005; Koblmüller et al. 2007). A complete list of the species sampled and their dominant diet is presented in Table 3-1.

Real-Time Quantitative Polymerase Chain Reaction

We measured opsin gene expression in each cichlid via real-time quantitative polymerase chain reaction (RT-qPCR). Our methods for RNA extraction and subsequent RT-qPCR analysis generally followed those previously used to analyze opsin expression in cichlids from LM and LV (Spady et al. 2006; Carleton et al. 2008; Hofmann et al. 2009). Binding sites for the Taqman primers and probes used in these previous studies were sequenced for all seven cone opsins from one individual of each species. Primers used to generate these sequences are listed in Supplementary Table S3-1 [Appendix 2]. Many LT species had opsin sequences that perfectly matched the primers created previously for LM and LV cichlids. In these cases, we used the primers and probes from these previous studies. However, where

these sequences differed, we created new LT-specific primers and probes. These new primers and probes, along with the original LM primers used, are listed in Table 3-2. In all, we identified 15 unique primer/ probe combinations needed to match the different LT species sampled (Table S3-2 [Appendix 2]; see also Table 3-1). We performed all RT-qPCR reactions on a LightCycler 480 (Roche). We normalized all RT-qPCR reaction efficiencies against a construct of cichlid opsins specially developed for the normalization of cichlid opsin RT-qPCR (Spady et al. 2006). In some LT species, however, the primer/probe-binding region did not match the sequence of the normalization construct. For these species, we normalized reaction efficiencies against known concentrations of a relevant cDNA sample or a ~120 bp oligomer encoding the primer- and probe-binding site. As in previous studies (Spady et al. 2006; Carleton et al. 2008; Hofmann et al. 2009), our measurement of *RH2A* expression combined the genetically and functionally similar *RH2A*α and *RH2A*β opsins. We quantified opsin expression twice for all individuals and averaged the results. We then averaged individual results to obtain one final, species-specific mean and variance of opsin expression.

Predicting Maximal Retinal Sensitivity from Opsin Gene Expression

Cichlid cone opsins are expressed within the retina in two distinct cell types: single-cone photoreceptors and double-cone photoreceptors (Bowmaker 1995; Carleton 2009). We predicted the wavelength of maximal sensitivity (PS_{max}) of each species' single- and double- cone photoreceptors from the results of our RT-qPCR analysis. We used the results of these estimates to infer how retinal sensitivities may vary as a result of changes in opsin gene expression (Carleton et al. 2008; Hofmann et

al. 2009). These estimates provide a useful descriptive statistic for how multivariate shifts in opsin expression may alter spectral sensitivity, but they are not meant to imply that we find fish with cones that exhibit these exact λ_{max} values (although the results can be quite similar). Following previous studies (Carleton et al. 2008; Hofmann et al. 2009), we used the equation:

$$PS_{max,\,C} = \frac{\Sigma(f_i \lambda_i)}{\Sigma f_i}$$

to calculate the predicted maximal sensitivity (PS_{max}) of cichlid single- and double-cones, where C is either the single- or the double-cone photoreceptor, f_i is the percent expression of the i^{th} opsin out of the total, and λ_i is the corresponding peak absorbance of that opsin in *Oreochromis niloticus* (Spady et al. 2006). Based on comparison of the λ_{max} of each cichlid opsin with the λ_{max} of single- and double-cone photoreceptors from several cichlid species—both of them measured physiologically using microspectrophotometry (Carleton 2009)—we used the expression of the *SWS1*, *SWS2B*, and *SWS2A* opsins when estimating the PS_{max} of single-cones and *RH2B*, *RH2A*, and *LWS* when estimating the PS_{max} of double-cones. In previous analyses, we refer to the descriptive statistic PS_{max} as ‘‘predicted single-/double-cone λ_{max}’’ (Hofmann et al. 2009) [see Chapter 1] or simply ‘‘single-/double-cone λ_{max}’’ (Carleton et al. 2008).

Opsin Sequence Divergence

Our estimation of photoreceptor PS_{max} assumes that all species exhibit opsin-coding sequences that are functionally identical to those of *Oreochromis niloticus*. This assumption is generally supported by microspectrophotometry results that

demonstrate little variation in the spectral absorption of cones from different cichlid species that express the same opsin palette (Jordan et al. 2006) (see also Table 1 in Carleton (2009)). Additionally, several studies have generally found little variation in the protein-coding sequence of each opsin across several cichlid species. These studies include sequences from 4 LT cichlids (Halstenberg et al. 2005; Spady et al. 2005), 16 LM cichlids (Parry et al. 2005; Spady et al. 2005; Hofmann et al. 2009), and 12 LV cichlids (Carleton et al. 2005a; Hofmann et al. 2009). Although mutations within opsin-coding sequences can play an important role in cichlid spectral adaptation (Sugawara et al. 2005; Terai et al. 2006), these shifts are generally small (5–15 nm). However, in order to further test this assumption, we sequenced the three most variable cichlid opsins—*SWS1*, *RH2Aβ*, and *LWS*—in a sub set of the LT species sampled. We then compared the coding regions of these opsins with those from *O. niloticus*. The primers used to sequence these opsins are listed in Table S3-1 [Appendix 2]. This analysis also provides an important estimate of the contribution that protein-coding mutations make to the evolution of spectral sensitivity in LT cichlids.

Phylogenetic Analysis

For our comparative analyses of opsin expression evolution, we reconstructed the phylogenetic relationships among the LT species sampled using three mitochondrial loci, *ND2* (1047 bp), *CYTB* (401 bp), and the D-loop (364 bp). These sequences were accessed through GenBank or else sequenced directly using previously published primers and protocols (Meyer and Wilson 1990; Taberlet et al. 1992; Kocher et al. 1995; Lee et al. 1995). Table 3-1 lists the accession numbers of

these sequences for each species, and Table S3-1 [Appendix 2] lists the primers used for PCR. Sequences were concatenated and aligned in MAFFT (Katoh et al. 2002), and we used Modeltest v3.7 (Posada and Crandall 1998) to choose an appropriate model of sequence evolution for the alignment. Phylogenetic reconstruction was performed using both Bayesian inference (BI) and maximum likelihood (ML) methods in the programs MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001) and PAUP v4.0b10 (Swofford 2003). For the BI analysis, the best-fit model of sequence evolution chosen by Modeltest (general time reversible GTR + Γ + I) was used to construct and run four Markov chain Monte Carlo (MCMC) sampling chains, each run for 1,000,000 generations with a swap frequency of once every 10 generations. Trees were sampled every 1,000 generations after discarding the first 10% as burn-in. We additionally discarded as burn-in the first 250 trees when calculating posterior probability values for the final 50% majority-rule consensus tree. For the ML analysis, we used the heuristic tree search method with random addition of sequences and tree bisection and reconnection branch swapping. In addition to the GTR + Γ + I model, for this analysis we also specified several additional model parameters estimated by Modeltest. These parameters were base frequencies (A = 0.2862, C = 0.3267, G = 0.1140, T = 0.2731), substitution rates (A–C = 0.7613, A–G = 11.8437, A–T = 1.2821, C–G = 0.6842, C–T = 6.5861, G–T = 1.000), proportion of invariable sites (0.4546), and the gamma distribution shape (0.9362). We used 100 ML bootstrap replicates to calculate nodal support for the final 50% majority-rule consensus tree. We rooted this tree with sequences from *O. niloticus* (AB018974, AF550020, and AF328847). We used this tree in all comparative analyses of opsin

expression with diet and lens transmittance among LT cichlids.

For the analyses of repeated evolution among cichlids from different East African rift lakes, we combined this mitochondrial tree with four published amplified fragment length polymorphism (AFLP) phylogenies of LM and LV cichlids (Albertson et al. 1999; Allender et al. 2003; Seehausen et al. 2003; Kidd et al. 2006). Due to the young age of the LM and LV cichlid species flocks (1.0 MY and >0.2 MY, respectively), the interrelationships of these taxa can only be resolved with genome-wide scans of many AFLP or single nucleotide polymorphism genotypes; mitochondrial DNA is not sufficient to resolve the phylogenies of these two groups. However, the monophyly of the LM and LV radiations, the structure of the LT radiation, and the interrelationships among taxa between the three major lakes have all been confidently resolved using mitochondrial loci (Meyer et al. 1990; Kocher et al. 1993; Salzburger et al. 2002; Kocher 2004) (see also Figure 3-1). The overall structure of our combined tree is consistent with the purported relationships of taxa in these three lakes (Salzburger et al. 2002; Koblmüller et al. 2008). The inter-relationships among taxa from LM and LV reported here use only those nodes supported by ≥60% bootstrap support in their respective studies. We set all branch lengths of this composite phylogeny to one.

Analysis of Parallel Evolution

To test the hypothesis that the various opsin gene expression palettes have evolved repeatedly among African cichlids from different rift lakes, we used two methods. First, we performed multivariate Bayesian clustering (Fuentes and Casella 2009; Gopal et al. 2009) to statistically group the 28 LT cichlids sampled here with

65 additional species from LM and LV (Hofmann et al. 2009), as well as the Nile tilapia, *O. niloticus* (Carleton et al. 2008). Because the cichlid species flocks of LM and LV both form monophyletic groups, we do not expect them to more closely resemble cichlids from LT unless they have evolved similar patterns of opsin gene expression in parallel. Bayesian clustering groups observations not by a distance-based metric but by a Metropolis search algorithm that attempts to maximize the marginal probability of $(Y \mid \omega_k)$, where Y is a matrix of response variables (e.g., opsin expression values for each species) and ω is the partitioning of Y into a pre-specified number of k clusters. This method then tests the statistical significance of the resulting clusters using a Bayes factor to estimate the empirical posterior probability (PP) of the null hypothesis

H_0: No clusters ($k = 1$) *versus* H_1: k clusters.

In order to generate a frequentist probability value for this test, we performed a second search of the PP space under the null hypothesis in order to generate a null distribution of quantiles for these values; we then compared the final PP value with this distribution (Fuentes and Casella 2009). For our analysis, we specified the presence of $k = 3$ clusters, representing the three opsin gene expression palettes so far observed in African cichlids (Fernald and Liebman 1980; Carleton et al. 2000; Carleton et al. 2005a; Parry et al. 2005; Jordan et al. 2006; Carleton et al. 2008; Hofmann et al. 2009). However, we also performed this analysis with k equal to 4 and 5 clusters as well. We performed Bayesian clustering in the R package ‘‘bayesclust’’ (Gopal et al. 2009). We used 1,000,000 simulations to estimate both the optimal clustering of the taxa and the PP of the null hypothesis. We used 10,000

simulations when generating the null distribution of PP values.

Second, we reconstructed the ancestral state of each major cichlid tribe using both Bayesian and ML methods. Using a composite phylogeny of 47 cichlids from all three lakes, we first estimated the posterior probability that the ancestor of each tribe expressed the opsin palette represented by $k = 3$, 4, and 5 clusters in the program BayesTraits (Pagel et al. 2004; Pagel and Meade 2007). BayesTraits infers ancestral states using a reversible-jump (RJ) MCMC search algorithm. For this analysis, we specified a reversible-jump hyperprior derived from the exponential distribution but seeded from a uniform (uninformative) distribution of values ranging from 0 – 30. We also specified a rate deviation parameter equal to one. Together, these parameters produced acceptance rates of newly proposed values equal to ~24%, which is within the desired range for this type of analysis (Pagel and Meade 2007). We ran the RJ-MCMC for 20,020,000 generations, discarded the first 20,000 generations as burn-in, and sampled the chains every 300 generations. All reconstructions were performed using the ‘‘BayesMultiState’’ module with the ‘‘AddNode’’ command.

Finally, we also reconstructed the ancestral state of each cichlid tribe following a ML analysis of each opsin's expression value in the R package APE v2.5 (Paradis et al. 2004). This analysis allowed us to reconstruct the ancestral state of each opsin individually, without forcing a discrete cluster assignment to each species' palette or the reconstructed ancestral states. However, we note that continuous character state reconstructions have been shown to perform poorly over adaptive radiations (Schluter et al. 1997).

For all ancestral state reconstructions, we rooted our tree of African cichlids from LT, LM, and LV with the tilapine cichlid *O. niloticus*. Both physiological measurements of retinal sensitivity as well as predictions made from opsin expression values indicate that *O. niloticus* expresses the long-wavelength palette (Spady et al. 2006; Carleton et al. 2008). These physiological and opsin expression values are representative of all additional riverine outgroups for which spectral sensitivities have been measured, including the tilapine *Sarotherodon* and the distantly related Neotropical cichlids (Levine and MacNichol Jr. 1979; Spady et al. 2006; Carleton et al. 2008; Carleton 2009). Thus, we present *O. niloticus* as a representative member of Oreochromis and other outgroups to the cichlids we include here.

Comparative Analyses with Diet and Lens Transmittance

We tested the hypothesis of correlated evolution among opsin expression, diet, and lens transmittance using the phylogenetic comparative method (Felsenstein 1985). For the analysis of opsin expression with diet, we used phylogenetic analysis of variance (ANOVA) (Garland Jr. et al. 1993) to compare the mean expression of each opsin among species grouped into five foraging levels (Table 3-1). We implemented phylogenetic ANOVA in the programs PDSIMUL v2.0 (Garland Jr. et al. 1993) and PHYLOGR (Díaz-Uriarte and Garland Jr. 2007). We performed 1,000 simulations of each opsin variable across the LT phylogeny using a Brownian motion model of character evolution. These simulations were used to generate phylogenetically corrected null distributions of our test statistics for phylogenetic ANOVA. However, prior to performing these simulations, we first transformed several opsin variables to better meet the ANOVA assumptions of homogeneity of

variances and normality of errors. These transformations were performed using Box–Cox powers estimated in the R package ‘‘car’’ (Fox 2008) and are presented in Table 3-3. We added 1.5 as a constant to each observation in order to maintain the order of the means before transformation. Additionally, for the comparison of the *SWS2A* opsin, we transformed the branch lengths of the mitochondrial tree using Grafen's (1989) rho ($\rho = 0.1$) and excluded *Neolamprologus tretocephalus* as an outlier from this analysis. We ultimately compared the probability values from these analyses with the Bonferroni-corrected significance threshold for 10 comparisons ($\alpha = 0.05/10$ comparisons = 0.005; see Table 3-3). Finally, for each opsin, we also estimated Pagel's (1999) λ via ML in the R package ‘‘geiger’’ (Harmon et al. 2009). Pagel's λ provides an important measure of association between the phylogeny and variance for a given trait.

For the analysis of opsin expression with lens transmittance, we extracted lenses for approximately half of the species sampled. We measured the transmission of these lenses using an Ocean Optics USB4000 spectrometer and a pulsed Xenon lamp (PX2, Ocean Optics). Our measurements followed the previously published protocols of Siebeck and Marshall (Siebeck and Marshall 2001). Transmission values were normalized to 1 at 600 nm and used to determine the wavelength of 50% transmission (T50). Because light must first pass through the lens before reaching the retina, lens transmittance can limit the wavelengths of light reaching the photoreceptors. This is particularly true for wavelengths at the short-wavelength end of the visible light spectrum (Siebeck and Marshall 2001). Because short-wavelength sensitivity is mediated by the single-cone photoreceptors in cichlids (Fernald and

Liebman 1980; Jordan et al. 2006; Carleton 2009), we tested the hypothesis of correlated evolution between lens T50 and the predicted maximal sensitivity (PS_{max}) of cichlid single-cones (see equation above). For this analysis, we used phylogenetically independent contrasts (PICs; (Felsenstein 1985)) implemented in the PDAP:PDTREE module (Midford et al. 2003) of the program Mesquite v1.12 (Maddison and Maddison 2001). We set all branch lengths to one and log transformed single-cone λ_{max} values to meet the assumptions of the independent contrasts method and normality of errors.

Results and Discussion

Tanganyikan Opsin Expression Diversity

Figure 3-2 illustrates the results of our RT-qPCR analysis for the 28 LT cichlids sampled. The expression values measured for each opsin ranged from 0% to 73% of total opsin expression (Table S3-3 [Appendix 2]). Despite previous analyses that reveal small but statistically significant differences in opsin expression between retinas extracted while in the field and those extracted after rearing for one generation in a laboratory setting (Hofmann et al. 2010b), we found no discernable differences between the retinas of wild-caught Tanganyikan cichlids processed in the field and those shipped to our laboratory (data not shown). The majority of species simultaneously expressed 3 or 4 of the 6 opsins measured. Importantly, these expression patterns generally matched those observed among cichlids from LM and LV, which are *SWS1-RH2B-RH2A* (short-wavelength sensitive), *SWS2B-RH2B-RH2A* (middle-wavelength sensitive), and *SWS2A-RH2A-LWS* (long-wavelength sensitive)

(Hofmann et al. 2009). However, many species also expressed appreciable amounts (between 5 – 18%) of a fourth opsin—typically *LWS*—which is also observed among some LM and LV species (Hofmann et al. 2009). In general, most of the species we sampled expressed opsins from either the middle- or the long-wavelength palettes. But, in contrast to many LM cichlids that express the long-wavelength palette, LT cichlids with this palette generally expressed *SWS2B* in place of *SWS2A*. Also in contrast to LM and LV cichlids, a few species expressed high levels of either *SWS1* or *SWS2A* (Figure 3-2). Finally, at least two species exhibited opsin expression profiles that had not been previously described in cichlids from LM and LV. *Paracyprichromis nigrapinnis* expressed high levels of *SWS1* in conjunction with *RH2A* and *LWS*, and *N. tretocephalus* expressed high levels of *SWS2A* in conjunction with *RH2B* and *RH2A* (Figure 3-2).

The approximate spectral sensitivity estimated for cichlids with these opsin expression palettes is illustrated in Figure 3-2. The average predicted maximal sensitivity (PS_{max}) for single-cones ranged nearly 100 nm, from 366 – 453 nm. The average joint double–cone PS_{max} for these species had a slightly narrower range, from 487 – 552 nm (Table S3-3 [Appendix 2]). The distribution of LT cichlids across the combined predicted sensitivity of these two photoreceptors reveals that LT cichlids likely exhibit spectral sensitivities that overlap those observed or predicted for cichlids from LM and LV (Figure 3-2). Among taxa with the middle-wavelength palette, several LT species also exhibited opsin expression profiles that were subtly divergent from those previously observed. Members of the tribe Lamprologini, including *Julidochromis regani*, *Neolamprologus brichardi*, *N. furcifer*, and

Chalinochromis brichardi, exhibited single-cone PS_{max} that were short-wavelength shifted relative to other species with this palette, and members of the tribe Ectodini, including *Enantiopus melanogenys* and X*enotilapia ochrog*enys, exhibited double-cone PS_{max} that were long-wavelength shifted (Figure 3-2). The novel opsin expression palettes of *P. nigrapinnis* and *N. tretocephalus* were predicted to confer visual pigment sensitivities with pigment spacings that were broader and narrower, respectively, compared with the three more common palettes. Once again, the results of our analysis of estimated photoreceptor sensitivities are not meant to imply that these species have photoreceptors with these exact absorbance values; rather they provide a useful summary statistic for estimating how multivariate changes in opsin gene expression may shift spectral sensitivity. However, the results of our opsin expression and photoreceptor PS_{max} analyses both suggest that visual system diversity is similar among African cichlids in LT and LM but that this diversity is potentially greater among the more phenotypically and phylogenetically diverse LT cichlids (Salzburger et al. 2002).

Opsin Sequence Diversity

Our analysis of opsin-coding sequences supports the assumption that LT cichlids possess opsins with λ_{max} similar to those of *O. niloticus*. We sequenced the *SWS1* opsin in 10 species and found that it was the most variable of the three opsins examined (Table S3-4 [Appendix 2]). We identified 25 polymorphic amino acid sites among these taxa; however, only six of these sites occurred in regions likely to affect chromophore binding and, therefore, spectral sensitivity. Of these 6 sites, 5 exhibited replacements that considerably alter the physical or chemical properties of the amino

acid substituted (A52T, A97S, I290T/S, and A298S). However, only one substitution (I290T/S) was absolutely fixed between *O. niloticus* and the LT cichlids. Substitutions at the remaining sites were shared between O. *niloticus* and other species, and no substitutions were found in sites already known to influence *SWS1* absorption (Yokoyama 2008). We then sequenced *RH2Aβ* in 14 species. Here we found 18 polymorphic sites, but only one of which occurred in a chromophore-binding region (Table S3-4 [Appendix 2]). This polymorphism, F203Y, varies in amino acid polarity but has not yet been shown to impact spectral tuning. However, it is possible that such a polarity shift could slightly impact the spectral absorption of the *RH2Aβ* opsin (Chang et al. 1995). Finally, we sequenced *LWS* in 11 species and found 10 variable sites. But, once again, we found only one site that occurred in a chromophore-binding region (Table S3-4 [Appendix 2]). This polymorphism, A164S, does change the amino acid polarity and has been shown to cause a 7 nm increase in *LWS* absorbance in humans (Asenjo et al. 1994) and LV cichlids (Terai et al. 2006). In summary, we found only one polymorphism in each opsin that was likely to produce a shift in the sensitivity of that gene relative to *O. niloticus*. Therefore, we conclude that LT cichlids have opsins with spectral sensitivities similar to those of *O. niloticus*, which justifies our use of *O. niloticus* opsin λ_{max} in the estimation of photoreceptor sensitivities. We emphasize that the sequence differences we observe would only produce small shifts (5–15 nm; (Carleton 2009; Hofmann et al. 2009)) in spectral sensitivity relative to the large shifts (30–100 nm) caused by changes in opsin gene expression. Therefore, none of the sequence substitutions we observe would alter the placement of LT species into different visual palette groups.

Interestingly, several of the sites we identified as polymorphic in the opsins of LT cichlids are also polymorphic among cichlids from LM and LV (e.g., *SWS1* site 217; *RH2Aβ* sites 107, 151, and 218; and LWS site 264) (Table S3-4 [Appendix 2]) (Hofmann et al. 2009). These mutations could indicate parallel mutations within opsin-coding sequences but could also be ancestral polymorphisms (Spady et al. 2005; Terai et al. 2006). We also found that the *SWS1* opsin exhibited the largest number of putatively functional replacements among cichlids from LT. Although we did not examine all opsins, this pattern is consistent with a specific role for protein-coding evolution within opsins sensitive to the ends of the visible light spectrum (Hofmann et al. 2009). However, only one site, *SWS1*-217, was variable in both these groups; the rest were unique to cichlids from each lake. This observation could suggest that there has been convergent functional evolution of the *SWS1* opsin in cichlids from LT and LM. This pattern is likely not the result of the rapid accumulation of deleterious alleles because none of the sequences we examined were pseudogenes, although we acknowledge that the *SWS1* opsin is not highly expressed among the adults of most LT species examined (Figure 3-2A). However, lack of *SWS1* expression among adults does not rule out its use earlier during development (e.g., Carleton et al. (2008)). The small number of putatively functional substitutions we identify in the remaining two opsins suggests that opsin protein-coding mutations likely contribute very little to divergence of spectral sensitivity among LT cichlids, with the possible exception of the *SWS1* opsin. However, we note that even small shifts in spectral sensitivity can impact female choice and even speciation (Seehausen et al. 2008). The much larger shifts in spectral sensitivity associated with changes in

opsin expression could have an even greater impact on divergence among cichlids from LM and LT.

Phylogenetic Analysis

The final 50% majority-rule consensus trees produced by our Bayesian and ML analyses were highly resolved and widely congruent. In each case, the positions of the major tribes were identical, and the trees differed only slightly in their branch lengths and support for certain nodes. Figure 3-2A illustrates the final consensus tree for both analyses incorporating BI-estimated branch lengths. Despite weak support for five nodes, our phylogeny is highly concordant with those previously reported for these or closely related species (Salzburger et al. 2002; Duftner, Koblmuller, Sturmbauer 2005; Day, Santini, Garcia-Moreno 2007).

Parallel Evolution of Opsin Gene Expression

The results of our Bayesian cluster analyses using $k = 3$, 4, and 5 clusters generated clustering schemes with empirical posterior probabilities (PP) equal to 1.07e-21, 2.35e-19, and 3.05e-16, respectively. All clustering schemes produced PP values that were a statistically better fit to the observed data than the null hypothesis of $k = 1$ clusters or no differences between species ($P < 0.0001$ in all three cases). Unfortunately, the implementation of Bayesian clustering we use here cannot currently estimate the optimal number of k clusters, and the PP values of different tests cannot be compared for this purpose because each PP is unique to the model of k clusters specified (Fuentes and Casella 2009). However, because previous estimates support $k = 3$ as the optimal number of opsin expression clusters among LM cichlids

(Hofmann et al. 2009), we were primarily concerned with the results of this analysis.

The clustering scheme for $k = 3$ grouped cichlids from LT, LM, and LV into clusters that chiefly reflected the three opsin expression palettes previously identified in these fishes (Table S3-5 [Appendix 2]). Group 1 consisted of three LT cichlids that express the short-wavelength opsin palette (*Eretmodus cyanostictus*, *Tanganicodus irsacae*, and *P. nigrapinnis*) as well as most members of the LM rock-dwelling (*mbuna*) lineage and also some members of the LM sand-dwelling (*utaka*) lineage (Figure 3-3; Table S3-5 [Appendix 2]). Group 2 consisted of taxa that express the middle-wavelength palette and included the majority of the LT cichlids sampled. Species in this group include members of the tribes Neolamprologini and Ectodini, *Benthochromis tricoti*, *Cyprichromis leptura*, *Petrochromis famula*, *Greenwoodichromis christyi*, and several members of both the LM *mbuna* and *utaka* lineages. Finally, group 3 consisted of seven LT cichlids that express the long-wavelength opsin palette, including *A. burtoni*, *Ophthalmotilapia ventralis*, *Neolamprologus tretocephalus*, most members of the tribe Tropheini, many members of the monophyletic LM *utaka* lineage, and all members of the monophyletic LV cichlid species flock. The results for $k = 4$ and 5 simply subdivided the short- and long-wavelength–sensitive clusters, respectively (Figure S3-1 and Table S3-5 [Appendix 2]). Grouping taxa into $k = 4$ clusters split species that express the two short- and middle-wavelength palettes into a third group of species that exhibit additional *SWS2B* and *LWS* expression. Grouping taxa into $k = 5$ clusters split taxa that express the long-wavelength palette into two groups based on those with additional *SWS2B* expression (Table S3-3 [Appendix 2]; see also Table S1-1

[Appendix 1]). Finally, we note that the clustering results of Bayesian clustering of k = 3 groups are very similar to the clustering scheme identified by principle component analysis and k-means clustering (data not shown). Thus, our results are robust to the analytical method used to group individuals based on opsin gene expression. In all cases, the statistically significant clustering of species from different, monophyletic lineages within LT, LM, and LV strongly supports the repeated evolution of opsin gene expression among African cichlids.

Reconstruction of the evolutionary history of these clusters on the phylogeny of African cichlids also supports the repeated evolution of multiple opsin expression palettes. Figure 3-3 illustrates the posterior probability that the ancestor of each major tribe expressed the palettes represented by k = 3 clusters following Bayesian ancestral state reconstruction. With only two exceptions, this reconstruction overwhelmingly supports the long-wavelength palette as the ancestral state for most major clades, including the haplochromine tribes of LM and LV. The two exceptions are the joint ancestor of the tribes Cyprichromini, Benthochromini, Perissodini, and Limnochromini, which likely expressed the middle-wavelength palette, and the ancestor of the Eretmodini, which expressed the short-wavelength palette. This reconstruction therefore indicates several transitions to the short- and middle-wavelength palettes among members of the various African cichlid lineages in LT and LM. Specifically, the short-wavelength palette arose twice within LT and then again among members of the LM cichlid radiation; the middle-wavelength palette arose at least four times among cichlids from LT and at least twice among cichlids from LM; and, because the long-wavelength palette is ancestral to most tribes, its

evolution does not appear to have occurred in parallel among the cichlids from LT, LM, and LV. However, this palette may have re-evolved at least once within the LM *utaka* clade (Figure 3-3). Reconstructions of $k = 4$ and 5 clusters on the cichlid phylogeny also overwhelmingly support a long-wavelength palette (group 3 or 5, colored red and yellow in Figure S3-1 [Appendix 2]) as the ancestral state for most African cichlid tribes. Reconstruction of $k = 5$ clusters suggests that the ancestors of each lineage gradually developed a violet- (as opposed to blue-) shifted long-wavelength palette leading up to the LV radiation. However, the reconstructions of both $k = 4$ and 5 clusters also indicate numerous transitions to the short- and middle-wavelength palettes among members of the LT tribes Eretmodini, Lamprologini, Ectodini, as well as LM cichlids (Figure S3-1 [Appendix 2]). Once again, these transitions occurred among species and lineages with different ancestors that each expressed the long-wavelength palette. Hence, even though we cannot distinguish between the optimality of $k = 3$, 4, or 5 clusters, the ancestral reconstruction of each of these scenarios all supports the parallel evolution of the short- and middle-wavelength palettes among cichlids from LT and LM from ancestors that expressed the long-wavelength palette.

Finally, we also used continuous character state reconstructions via ML to infer ancestral states of each opsin's expression pattern independently. This continuous character reconstruction produced estimates of ancestral states that were highly uncertain. Ninety-five percent confidence intervals for the inferred ancestral states overlapped for expression values at many nodes. Among internal nodes, only states at the base of the clades Eretmodini, Lamprologini, and Benthochromini

deviated significantly from the states of their direct ancestor along the base of the tree (indicated by pluses and minuses in Figure S3-1 [Appendix 2]). In contrast, many species (tips) had 95% confidence intervals that did deviate significantly from the expression values predicted for the ancestor at the base of their respective clade. This pattern could be due to the inherent uncertainty in the ancestral states of nodes further from the tips of the phylogeny; however, we believe this pattern indicates that most shifts in opsin expression have occurred near the tips of the cichlid phylogeny, not at its base. To account for this possible bias, we also identified shifts in opsin expression of greater than 10% (indicated by greater than and less than symbols in Figure S3-1 [Appendix 2]). This analysis illustrates the same pattern: few large shifts in expression at internal nodes, except for the base of the clades Eretmodini, Lamprologini, Benthochromini, and Tropheiini. Once again, most shifts in opsin expression of more than 10% were concentrated at the tips of the phylogeny, indicating that this observation is not merely the result of statistical uncertainty in the ancestral states of internal nodes. Most shifts in opsin expression that were statistically significant (e.g., where 95% confidence intervals between an ancestor and descendent node did not overlap) were ≥10%. Additionally, our analysis based on percent expression is necessary where multiple observations of a particular species are not available to generate confidence intervals, which was the case for most LM cichlids. However, analyses of both confidence intervals and percentage units infer many parallel shifts in opsin expression. Most shifts represent increases in expression of the *SWS1*, *RH2B*, and *RH2A* opsins among cichlids from LT and LM (Figure S3-1 [Appendix 2]). We also find evidence for many independent shifts to lower

expression levels for the *SWS2A* and *LWS* opsins. Both observations are consistent with the parallel evolution of the short (*SWS1*-*RH2B*-*RH2A*) and middle (*SWS2B*-*RH2B*-*RH2A*) palettes. Most importantly, the results of our continuous character state reconstructions are highly concordant with the results of our Bayesian reconstructions of $k = 3$ opsin palettes (Figure 3-3). By estimating the spectral sensitivity of each ancestor via estimated single- and double-cone PS_{max}, we demonstrate that most ancestors exhibit inferred opsin expression values consistent with the long-wavelength palette (inset in Figure S31 [Appendix 2]). The only nodes that deviate from the long-wavelength palette are nodes at the base of the clades Eretimodini and Benthochromini. This observation is perfectly consistent with our reconstruction of $k = 3$ clusters (Figure 3-3). Thus, we conclude that similar opsin expression profiles among cichlids from LT and LM are due to parallel evolution from ancestors that each expressed the long-wavelength–sensitive opsin palette.

We refer to the repeated evolution of similar opsin expression profiles among cichlids from LT and LM as due to parallelism because, in each case, these transitions have occurred independently among taxa with different ancestors that shared the same ancestral state. This pattern contrasts with the expectation due to convergence, where similar phenotypes evolve among unrelated taxa whose ancestors exhibited different ancestral states. (We note, however, that convergence and parallelism likely represent to ends of a continuum of homoplasy ((Arendt and Reznick 2008))). We cannot infer how the palettes evolved in parallel with our current data. One hypothesis is that the presence of similar opsin expression profiles among cichlids in LT and LM is simply due to the sorting of ancestral polymorphism that affects adult

variation in opsin expression. We do not believe this is the case because the presence of alternate opsin expression palettes has not been reported among the adults of any one cichlid population or species. This observation suggests that the ancestral groups likely did not exhibit this much population-level variation either. A second hypothesis is that these palettes evolved independently among an ancestral group of haplochromine cichlids that subsequently produced a hybrid swarm (e.g., Seehausen (2004)). These palettes could then have been sorted coincident with the formation of new species. This hypothesis would produce the appearance of ancestral polymorphism; however, it would still indicate that the short- and middle-wavelength palettes evolved in parallel among LT and LM cichlids, only with a much earlier origin than our current phylogeny suggests (near the base of the LM clade instead of near the tips). Both the sorting of ancestral polymorphisms and a hybrid swarm scenario are consistent with what has been shown for the evolution of pigmentation blotching in LM cichlids (Roberts, Ser, Kocher 2009) and mitochondrial loci (Moran and Kornfield 1993). However, we favor a third hypothesis that the presence of similar opsin expression profiles among unrelated cichlids in LT and LM is the result of parallel heterochronic shifts in opsin expression from ancestors that expressed the entire complement of opsin palettes during development (e.g., Carleton et al. (2008)). Both the basal cichlid *O. niloticus* (Carleton et al. 2008) and the derived haplochromine cichlid *A. burtoni* (O'Quin et al. 2011) [see Chapter 3] express the short- and middle-wavelength palettes as fry and juveniles, respectively, but then consistently express the long-wavelength palette as adults. We believe that the presence of ontogenetic variation in opsin expression among both basal and derived

cichlids indicates that intermediate ancestral species (e.g., nodes b–g in Figure S3-1 [Appendix 2]) which are predicted to express the long-wavelength palette, probably did so following a similar developmental progression. If this is indeed the case, the presence of similar short- and middle-wavelength palettes among cichlids in LT and LM would be due to independent, heterochronic shifts in opsin expression from these ancestors (Carleton et al. 2008; O'Quin et al. 2011). However, additional sampling of LT cichlids with all three palettes at different ontogenetic stages will be necessary to conclusively test this hypothesis. We note also that our results rely on the observation that basal riverine cichlids express the long-wavelength opsin palette, of which *O. niloticus* is representative. Although this appears to be the case for all known African and Neotropical outgroups so far surveyed (Levine, MacNichol Jr. 1979; Carleton et al. 2008), sampling of additional genera such as Tylochromis and Tilapia may strengthen this conclusion.

Comparative Analyses with Diet and Lens Transmittance

To determine whether diet is associated with opsin expression divergence in LT cichlids, we compared the mean expression of each opsin among LT species divided into five foraging groups (Table 3-1). The ML estimates of Pagel's (1999) λ for each opsin generally indicate a weak association with phylogeny for the *SWS1* opsin and little or no association for the remaining opsins (Table 3-3). Despite these weak associations, we still use appropriate phylogenetic comparative methods for all comparisons. In our overall phylogenetic ANOVA, we found no statistically significant association between diet and mean relative expression for any of the opsins examined (Table 3-3). However, we did identify a similar trend of increased

SWS1 expression among zooplanktivorous cichlids in both LT and LM (Figure 3-4). Among cichlids from LM, diet is an important predictor of mean *SWS1* expression as well as actual and predicted single-cone λ_{max} (Jordan et al. 2004; Hofmann et al. 2009). LM cichlids that forage on zooplankton, algae, and phytoplankton on average exhibit higher levels of *SWS1* expression than cichlids that forage on fish or benthic invertebrates (Hofmann et al. 2009) (Figure 3-4). *SWS1* expression increases sensitivity to ultraviolet (UV) light, which has been shown to increase the ability of teleost fish to detect and feed on zooplankton because the UV-absorbing zooplankton appear as dark objects against the bright UV background (Novales-Flamarique and Hawryshyn 1994). Therefore, this trend motivated us to perform a post hoc Dunnett's test contrasting mean *SWS1* expression among LT cichlids that forage on zooplankton with the remaining foraging groups (Table 3-3).

We found that mean *SWS1* expression was significantly higher among zooplanktivorous species versus benthivores (phylogenetic t-test; $t = 3.174$, $p = 0.004$); however, we found no difference in mean *SWS1* expression between zooplanktivores and the remaining foraging groups. This weak but interesting correlation suggests that similar associations between *SWS1* expression and diet may have evolved independently among cichlids from both LT and LM (see results of ancestral character state reconstruction). The evolution of the same phenotypic correlation among unrelated cichlids in LT and LM could implicate natural selection in the parallel evolution of opsin expression among these species (Schluter 2000), since this association is unlikely to evolve repeatedly by drift alone. Future studies of additional zooplanktivorous cichlids in LT may bolster this conclusion.

In addition to diet, we also examined the correlated evolution of single-cone PS_{max} with lens transmittance. Lens transmittance (T50) values from LT cichlids were continuously distributed and ranged from 348.5 – 409 nm (Table S3-4 [Appendix 2]). Lens T50 was positively correlated with predicted single-cone PS_{max} (PICs: $r^2 = 0.417$, $F_{1,11} = 6.717$, $p = 0.013$) (Figure 3-4). Additionally, lens transmittance wavelengths were always lower than predicted single-cone PS_{max}, except in the case of *P. nigrapinnis*. These results indicate that cichlid lenses generally do not block wavelengths of light that the fish are highly sensitive to. Among LM cichlids, lens transmittance is also positively correlated with relative *SWS1* expression and estimated single-cone PS_{max} (Hofmann et al. 2010a), although lens T50 values are more bimodally distributed among these species (Figure 3-4). Interestingly, we identified four LT cichlids with lens T50 values that are intermediate to the two broad groups found among LM cichlids (Figure 3-4). These species are *G. christyi*, *N. cunningtoni*, *O. ventralis*, and *P. famula*. All these species are from different tribes but express either the middle- or the long-wavelength palette. Additionally, all these species' opsin expression palettes generally overlap those observed in LM cichlids, suggesting that these intermediate lens transmittance values are not associated with novel or unusual patterns of opsin expression (Figure 3-2; Table S3-3 [Appendix 2]). Like the results of our analyses of opsin expression diversity and photoreceptor sensitivity, the lens T50 values we observe suggest that visual system diversity is greater among the phylogenetically and phenotypically diverse cichlids of LT. Even so, the presence of similar, positive correlations between opsin expression divergence (illustrated through average single-cone PS_{max})

and lens transmittance among cichlids from LT and LM again suggests a role for natural selection in the parallel evolution of these traits.

We find that diet and lens transmittance are both associated with the evolution of opsin expression in cichlids from LT and LM, as they are in other groups as well (Munz, McFarland 1977; Lythgoe 1979; Losey et al. 2003). However, these two factors alone cannot explain all the similarities and differences in opsin expression we observe among cichlids from these two lakes. To illustrate this point, we identified three LT cichlids that are ecologically or morphologically similar to species in LM (Kocher et al. 1993; Kassam et al. 2003). The first pair of species, *Petrotilapia famula* (LT) and *Petrochromis nigra* (LM), both graze on epilithic algae and possess parallel morphological adaptations for doing so (Kassam et al. 2003). *P. famula* (LT) expresses the middle-wavelength palette, whereas *P. nigra* (LM) expresses the short-wavelength palette (Hofmann et al. 2009). These taxa also exhibit lens T50 that differ by ~15 nm (Hofmann et al. 2009). Similarly, both *Lobochilotes labiatus* (LT) and *Placidochromis milomo* (LM) possess puffy, distended lips for sucking invertebrates from the surface of rocks (Kocher et al. 1993). But we find that *L. labiatus* (LT) expresses the long-wavelength palette, whereas *P. milomo* (LM) expresses the middle-wavelength palette (Hofmann et al. 2009). The lens T50 of these two species differ by >40 nm (Hofmann et al. 2009). Only the final comparison between *Julidochromis regani* (LT) and *Melanochromis auratus* (LM), which both feed on phytoplankton and algae and both express the middle-wavelength palette, supports the hypothesis of ecological as well as spectral convergence. Unfortunately, we do not have lens transmittance data for *J. regani*.

In addition to differences in foraging preference and opsin expression in these three species comparisons, we also found that LT cichlids that forage on phytoplankton exhibit levels of *SWS1* expression on par with species that forage on fish and benthic invertebrates. This pattern contrasts strongly with ecologically similar species from LM (Figure 3-4). This difference is likely due to the expression of the long-wavelength palette among members of the LT tribe Tropheinii, which are phytoplanktivorous. This and the other examples we detail above likely contributed to the weak conclusion of our phylogenetic ANOVA (Table 3-3). To us, these observations suggest that other factors must also drive opsin expression evolution in African cichlids. These factors likely include additional ecological factors such as depth, as well as nonadaptive factors such as random genetic drift.

One additional ecological factor that could also explain the parallel evolution of similar opsin expression profiles among cichlids from LT and LM is the ambient light environment. Changes in spectral sensitivity due to the attenuation of light at different depths are observed among cichlids from all three East African Great Lakes (Sugawara et al. 2005; Seehausen et al. 2008). However, we were unable to test for an association between opsin expression and ambient light environment because detailed spectral measurements for LT are not available. Additionally, we had limited information regarding the sampling depth for most species. However, we note that the amount of opsin expression diversity present among cichlids from each lake seems to be correlated with the amount of spectral variation present in each lake. In other words, both LT and LM are remarkably clear and have waters with similar spectral qualities (Carleton et al. 2006). Cichlids from both these lakes exhibit a

diverse range of opsin expression profiles (e.g., at least three; see Figures 2-2 and S2-1) that collectively confer sensitivity to the entire spectrum of visible light available (Hofmann et al. 2009). In contrast, LV has a spectrally narrow light environment that is red shifted relative to LT and LM (Seehausen et al. 1997; Carleton et al. 2006). Opsin expression diversity in LV is very limited (Figures 2-3 and S2-1) and appears to be constrained to only those opsins sensitive to the long wavelengths of light present in the lake (Hofmann et al. 2009). These observations suggest that ambient light may also influence the evolution of opsin gene expression in African cichlids; however, future spectral measurements of LT will be necessary to definitively test this hypothesis.

Conclusions

Repeated phenotypic evolution can provide valuable insights into which genetic mechanisms generally contribute to the evolution of phenotypic diversity. Like pelvic spine loss in sticklebacks (Chan et al. 2010) and wing pigmentation in Drosophila (Prud'homme et al. 2006), we infer that cichlids in LT and LM have independently evolved similar retinal sensitivities through the parallel evolution of opsin gene regulation (Figures 2-2 and 2-3). Multiple ancestral state reconstructions support the parallel evolution of two distinct opsin expression profiles among unrelated cichlids from these two lakes (Figures 2-3). In contrast, we identified few protein-coding mutations that were likely to shift cichlid retinal sensitivities, with the possible exception of the *SWS1* (UV) opsin (Table S3-4 [Appendix 2]). Although opsin genes provide a classic example of how mutations within the protein-coding

regions of genes can contribute to phenotypic evolution (Yokoyama 2002), the independent evolution of similar opsin expression palettes among African cichlids underscores the important contribution that regulatory mutations can also make (Britten, Davidson 1971; King, Wilson 1975; Sucena et al. 2003; Prud'homme et al. 2006).

Why changes in opsin expression are prominent among cichlids from LT and LM could be due to similar adaptations to diet and lens transmittance (Figure 3-4), the light environment, or all three. Alternatively, biases in the use of one mutational type versus another could be due to selection (Schluter 2000) or genetic and developmental constraints (Schluter 1996; West-Eberhard 2003). For example, regulatory mutations may have relatively higher fitness when large shifts in opsin expression are necessary for spectral adaptation. In contrast, protein-coding mutations may be better suited for fine-tuning spectral sensitivity and necessary for turning spectral sensitivity at the two ends of the visible light spectrum (Hofmann et al. 2009). Examples of convergence in cichlid opsin-coding sequences do exist, particularly in the *RH1*, or rod, opsin (Sugawara et al. 2005). However, the spectral sensitivity of the rod opsin can only evolve through protein-coding mutations in cichlids because they do not have an additional *RH1* opsin to express. But in teleosts that do possess more than one rod opsin, large shifts in dim-light spectral sensitivity are generated through changes to the regulation of these genes (Yokoyama et al. 2008).

Exactly how the parallel evolution of opsin expression has been achieved among African cichlids from LT and LM is unclear. Currently, we cannot distinguish

between the hypotheses of de novo mutation, sorting of ancestral polymorphism, or parallel heterochronic shifts in opsin expression, although we favor the latter hypothesis. We have recently demonstrated that adult opsin expression has a strong genetic basis and is heritable (Carleton et al. 2010; Hofmann et al. 2010a). Further, hybrid crosses reveal that as few as two loci may underlie these important differences, including both *cis*- and *trans*-acting loci (Carleton et al. 2010). Future work will aim to use these hybrid crosses to further elucidate the molecular genetic basis for differential opsin expression in cichlids. We will also examine diversity at important *cis*- and *trans*- regulatory regions to determine what contributions these two mechanisms make to the evolution of spectral sensitivity in cichlids. These future analyses will help us distinguish between the possible scenarios that led to the parallel evolution of opsin gene expression among African cichlids.

Tables

Table 3-1. Lake Tanganyika cichlid species used in Chapter 2.

Species	*n*	Tribe[1]	Foaging	ND2	CYTB	D-LOOP	qPCR Comb.[h]
Benthochromis tricoti	3	Benthochromini	Benthic Invertebrates[2]	AY682515	AF428164	AY682477	15
Cyprichromis leptosoma	5	Cyprichromini	Zooplankton[2, 3]	AY740343	AB280682	AY740320	3
Paracyprichromis nigrapinnis	3	Cyprichromini	Zooplankton[2, 3]	AY740339	AY740204	AY740282	4
Asprotilapia leptura	2	Ectodini	Epilithic Algae[2]	AY337772	AY337801	AF400701	5
Enantiopus melanogenys	3	Ectodini	Benthic Invertebrates[2]	AY682517	AY337813	AY682480	2
Ophthalmotilapia ventralis	3	Ectodini	Zooplankton[2]	AY337774	AY337805	AY615479	14
Xenotilapia bathyphila	3	Ectodini	Benthic Invertebrates[2]	AY337789	AY337844	AY339027	2
Xenotilapia boulengeri	3	Ectodini	Benthic Invertebrates[2]	HM135111	AY337823	AY339029	10
Xenotilapia flavipinnis	2	Ectodini	Benthic Invertebrates[2]	AY337794	AY337825	AY339030	2
Xenotilapia ochrogenys	4	Ectodini	Benthic Invertebrates[2]	AY337767	Z21772	Z21750	2
Xenotilapia spiloptera	3	Ectodini	Benthic Invertebrates[2]	AY337788	AY337841	AY339040	2
Eretmodus cyanostictus	4	Eretmodini	Epilithic Algae[2]	AF398220	Z97477	EF035326	4
Tanganicodus irsacae	3	Eretmodini	Epilithic Algae[2, 4]	AF398219	Z97557	Y15134	3
Astatotilapia burtoni	1	Haplochromini	Benthic Invertebrates[2]	AF317266	Z21773	Z21751	5
Chalinochromis brichardi	2	Lamprologini	Phytoplankton[2]	HM135112	Z29991	Z30006	11
Julidochromis regani	2	Lamprologini	Phytoplankton[2, 5]	EF462228	EF470898	U01106	6
Neolamprologus brichardi	4	Lamprologini	Zooplankton[2]	DQ055015	AF438804	Z30021	7
Neolamprologus cunningtoni	4	Lamprologini	Fish[2]	HM135113	HM135105	HM135109	3
Neolamprologus furcifer	3	Lamprologini	Benthic Invertebrates[2, 6]	EF462249	Z29999	Z30026	5

Continued

Table 3-1 (continued).

Species	*n*	Tribe[1]	Foaging	ND2	CYTB	D-LOOP	qPCR Comb.[h]
Neolamprologus mondabu	3	Lamprologini	Benthic Invertebrates[2, 6]	EF462242	HM135106	HM135110	8
Neolamprologus tretocephalus	3	Lamprologini	Benthic Invertebrates[2, 6]	DQ055026	HM135107	-	13
Greenwoodichromis christyi	5	Limnochromini	Benthic Invertebrates[3]	AY682528	HM135108	AY682489	12
Perissodus microlepis	2	Perissodini	Fish[2, 7]	DQ055006	AF428167	EF437536	9
Lobochilotes labiatus	1	Tropheini	Benthic Invertebrates[2]	U07254	AY301932	U01110	1
Petrochromis famula	2	Tropheini	Epilithic Algae[2]	HM135114	AY301937	AY301963	1
Simochromis diagramma	5	Tropheini	Phytoplankton[2]	AY930087	AY301951	AY574628	1
Tropheus moori 'muzi'	4	Tropheini	Epilithic Algae[2]	AB018975	AB018990	Z12069	1
Tropheus sp. mpimbwe	3	Tropheini	Epilithic Algae[2]	AY930086	EF470900	Z12054	1

[1] Takahasi (2003)
[2] Brichard (1989)
[3] Duftner et al. (2005)
[4] Yamaoka et al. (1986)
[5] Salzburger et al. (2002)
[6] Taborsky et al. (1986)
[7] Koblmüller et al. (2007)
[8] Primer/probe combination used for RT-qPCR (see Supplementary Table S3-2 [Appendix 2]).

Table 3-2. Sequence of all primers and probes used to measure cichlid opsin gene expression.

Opsin	Primer	Sequence
SWS1	UV.Cic.Forward[1]	5'-GGCTGTGCCTGCCCAC-3'
	UV.Tang.Forward[2]	5'-GGCTGCGCCTGCCCAC-3'
	UV.Tang.Ov.Forward[2]	5'-TGCTGCGCCTTCCCAC-3'
	UV.Cic.Reverse[1]	5'-AGGAGCAGCCCAGACCTTC-3'
	UV.Cic.Probe[1]	5'-TTTCTTTGGCTGGAGCAGGTACATCCC-3'
SWS2B	B2.Cic.Forward[1]	5'-TTTGGTGCGCTAGCATGC-3'
	B2.Cic.Reverse[1]	5'-AAGGGACCACAGGCTTACCAT-3'
	B2.Cic.Probe[1]	5'-AGATCGAAGGTTTCATGGTAACACTCGGTG-3'
SWS2A	B1.Cic.Forward[1]	5'-TTTGGTGCGCTAGCATGC-3'
	B1.Tang.Reverse[2]	5'-CTTGCAAATCACAAGCCATC-3'
	B1.Cic.Probe[1]	5'-AGATCGAAGGTTTCATGGTAACACTCGGTG-3'
	B1.Tang.Probe[2]	5'-AGATCGAAGGTTTCATGGCAACACTCGGTG-3'
	B1.Tang.Nb.Probe[2]	5'-AGATCGAAGGTTTCATGGCAACACTTGGTG-3'
	B1.Tang.Nm.Probe[2]	5'-AGTTCGAAGGTTTCATGGCAACACTCTGTG-3'
	B1.Tang.Pm.Probe[2]	5'-AAATCGAAGGTTTCATGGCAACACTCGGTG-3'
	B1.Tang.Xeno.Probe[2]	5'-AGATCGAAGGTTTCTTGGCAACACTCGGTG-3'
RH2B	G3.Cic.Forward[1]	5'-TGCTGCCCCCCCATTG-3'
	G3.Cic.Reverse[1]	5'-AGGTCCACAGGAAACCTGAA-3'
	G3.Cic.Probe[1]	5'-TGGCTGGTCAAGGTACATTCCTGAGGGA-3'
RH2A	G.Tang.Forward[2]	5'-TTAATGGCTACTTCATTCTTGGA-3'
	G.Cic.Reverse[1]	5'-CCAGGACAACAAGTGACCAGAG-3'
	G.Cic.Probe[1]	5'-TGGCCACACTAGGAGGTGAAGTTGC-3'
	G.Til.Probe[1]	5'-TGGCCACACTTGGAGGTGAAGTTGC-3'
	G.Tang.Gc.Probe[2]	5'-TGGCCACACTTGGAGGTGAAGTTTC-3'
	G.Tang.Ov.Probe[2]	5'-TGGCCACACTAGGAGGTCAAGTTGC-3'
LWS	R.Cic.Forward[1]	5'-CTGTGCTACCTTGCTGTGTGG-3'
	R.Cic.Reverse[1]	5'-GCCTTCTGGGTTGACTCTGACT-3'
	R.Tang.Nb.Reverse[2]	5'-GCTTTCTGGGTTGACTCTGACT-3'
	R.Tang.Nt.Reverse[2]	5'-GCCTTTTGGGTTGACTCTGACT-3'
	R.Tang.Xb.Reverse[2]	5'-GCCTTCTGGGTTGACTCTGATT-3'
	R.Cic.Probe[1]	5'-TGGCCATCCGTGCTGTTGCC-3'

[1] Spady et al. (2006)
[2] This study

Table 3-3. Results of phylogenetic ANOVA comparing opsin gene expression with foraging preference and *post hoc* comparisons of *SWS1* expression between foraging levels.

Opsin	Pagel's λ	Box-Cox Power	df	*F* or *t* value	p-value
SWS1	0.3988	-1.772	4, 23	2.587	0.099
Zooplankton vs. Epilithic Algae	-	-	1, 23	1.823	0.115
Zooplankton vs. Phytoplankton	-	-	1, 23	1.945	0.071
Zooplankton vs. Fish	-	-	1, 23	1.462	0.129
Zooplankton vs. Benthic Invertebrates	-	-	1, 23	3.174	0.004**
SWS2B	0.1644	-	4, 23	1.465	0.337
SWS2A	<0.0001	-1.541	4, 22	0.541	0.820
RH2B	0.1301	-	4, 23	0.959	0.557
RH2A	<0.0001	0.188	4, 23	0.209	0.952
LWS	0.1796	0.350	4, 23	0.989	0.528

** These tests are significant following Bonferonni-correction for 10 hypothesis tests ($\alpha = 0.05/10 = 0.005$).

Figures

Figure 3-1. Schematic of the East African Great Lakes and the phylogenetic structure of their associated cichlid species flocks. (A) Map of the African continent with the location of the three Great Lakes – LT, LM, and LV—shown in gray. (B) Representative phylogeny of cichlids from each of the Great Lakes, with approximate dates of divergence (modified from Kocher (2004) and Koblmüller et al. (2008)). Map modified from the R package "maps" (Becker et al. 2010).

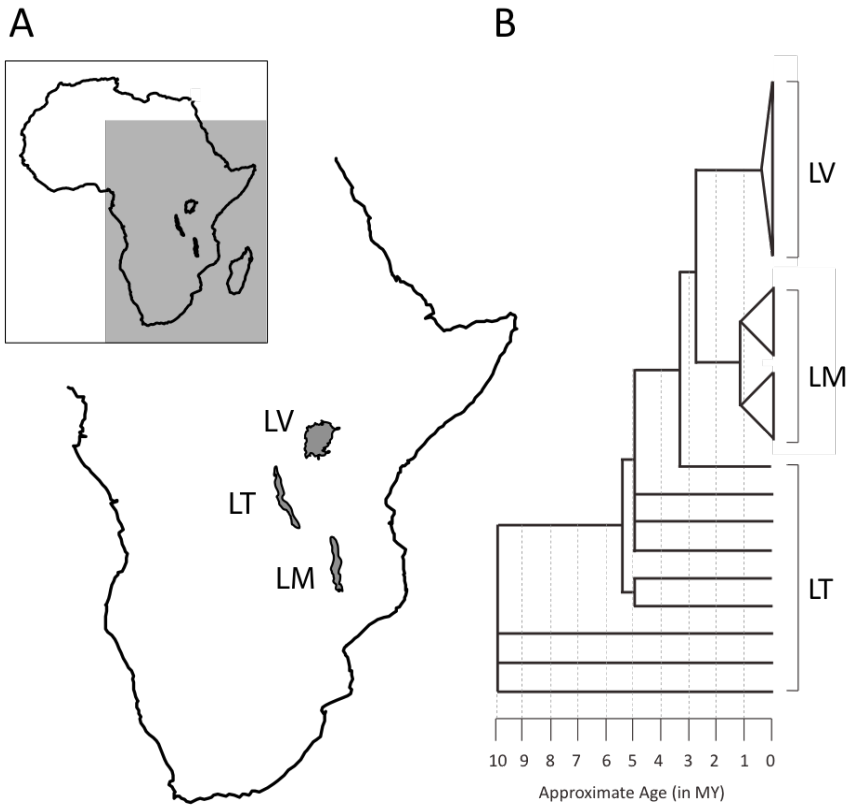


Figure 3-2. Opsin expression diversity in 28 cichlid species from LT. (A) Mitochondrial phylogeny of the species sampled. Filled circles indicate notes with > 80% bootstrap and posterior probability support; gray circles, nodes with > 50% bootstrap and posterior probability support; open circles, nodes with > posterior probability support only. (B) Heat map of relative opsin gene expression. The tribe to which each species belongs is shown on the left along with the visual palette estimated from the opsin expression profile (see text). (C) Predicted maximal sensitivity (PS_{max}) of single- and double-cone photoreceptors estimates from the opsin expression results. The distribution of photoreceptor sensitivities estimates for cichlids from LM and LV are indicated by gray boxes, including those expressing the short- (S), middle- (M), and long- (L) wavelength sensitive opsin palettes (Hofmann et al. 2009) [see Chapter 2].

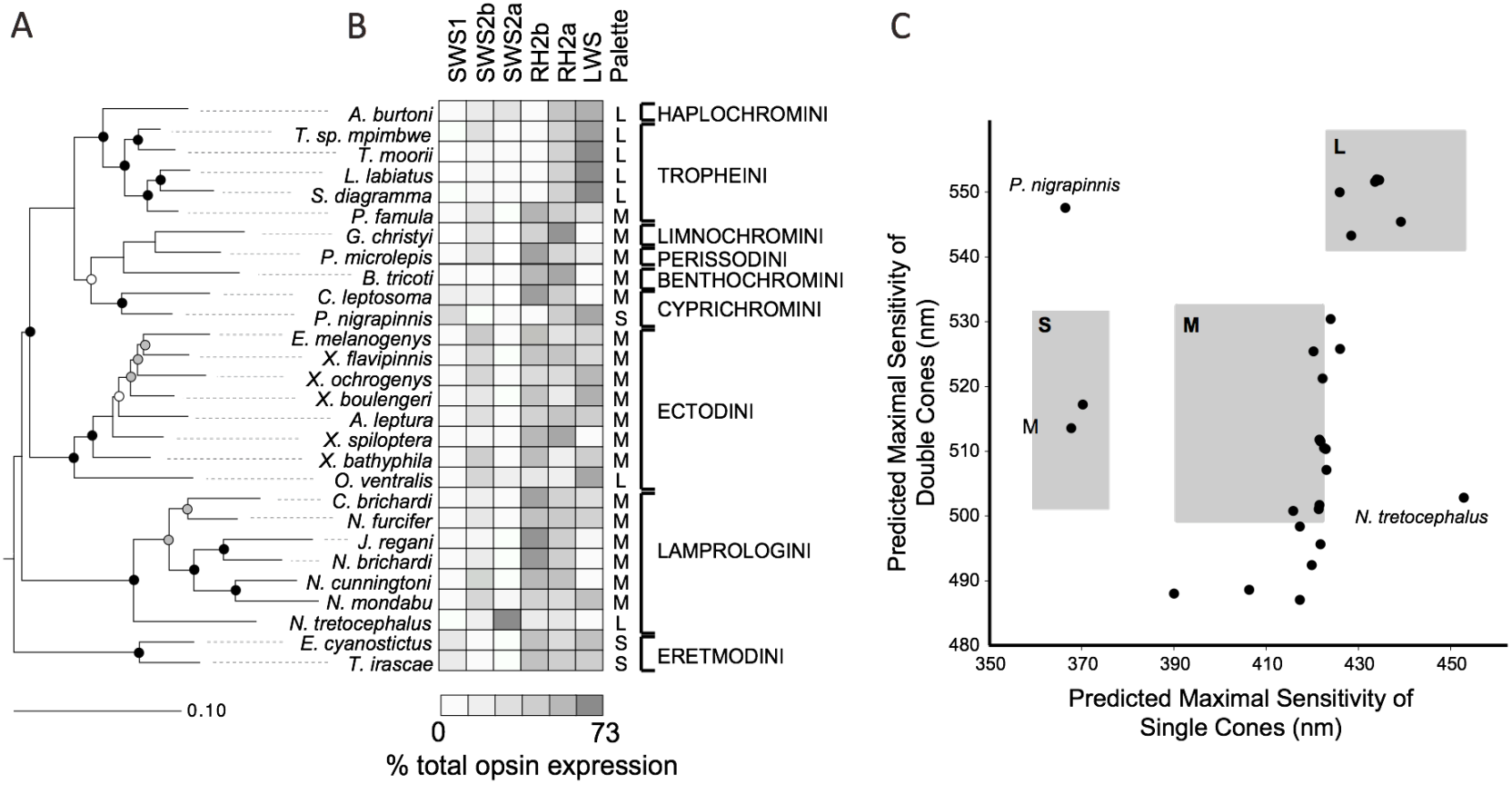

A
B
C
SWS1
SWS2b
SWS2a
RH2b
RH2a
LWS
Palette
A. burtoni
T. sp. mpimbwe
T. moorii
L. labiatus
S. diagramma
P. famula
G. christyi
P. microlepis
B. tricoti
C. leptosoma
P. nigrapinnis
E. melanogenys
X. flavipinnis
X. ochrogenys
X. boulengeri
A. leptura
X. spiloptera
X. bathyphila
O. ventralis
C. brichardi
N. furcifer
J. regani
N. brichardi
N. cunningtoni
N. mondabu
N. tretocephalus
E. cyanostictus
T. irascae
HAPLOCHROMINI
TROPHEINI
LIMNOCHROMINI
PERISSODINI
BENTHOCHROMINI
CYPRICHROMINI
ECTODINI
LAMPROLOGINI
ERETMODINI
0.10
0
73
% total opsin expression
Predicted Maximal Sensitivity of Double Cones (nm)
Predicted Maximal Sensitivity of Single Cones (nm)
P. nigrapinnis
N. tretocephalus
L
M
S
550
540
530
520
510
500
490
480
350
370
390
410
430
450

Figure 3-3. Parallel evolution of opsin gene expression in 47 African cichlid fishes from LT, LM, and LV, as well as the rivers (R). Pie charts illustrate the results of Bayesian ancestral state reconstruction and show the relative posterior probability that the ancestor expressed each of the three opsin expression palettes determined by clustering taxa into $k = 3$ clusters. The long-wavelength (red) palette is supported as the ancestral state for most African cichlid lineages, including the Haplochromini (LM and LV). States at the tips indicate several parallel shifts to the short- (blue) and middle- (green) wavelength palettes among cichlids in LT and LM from ancestors that each expressed the long-wavelength palette (red).

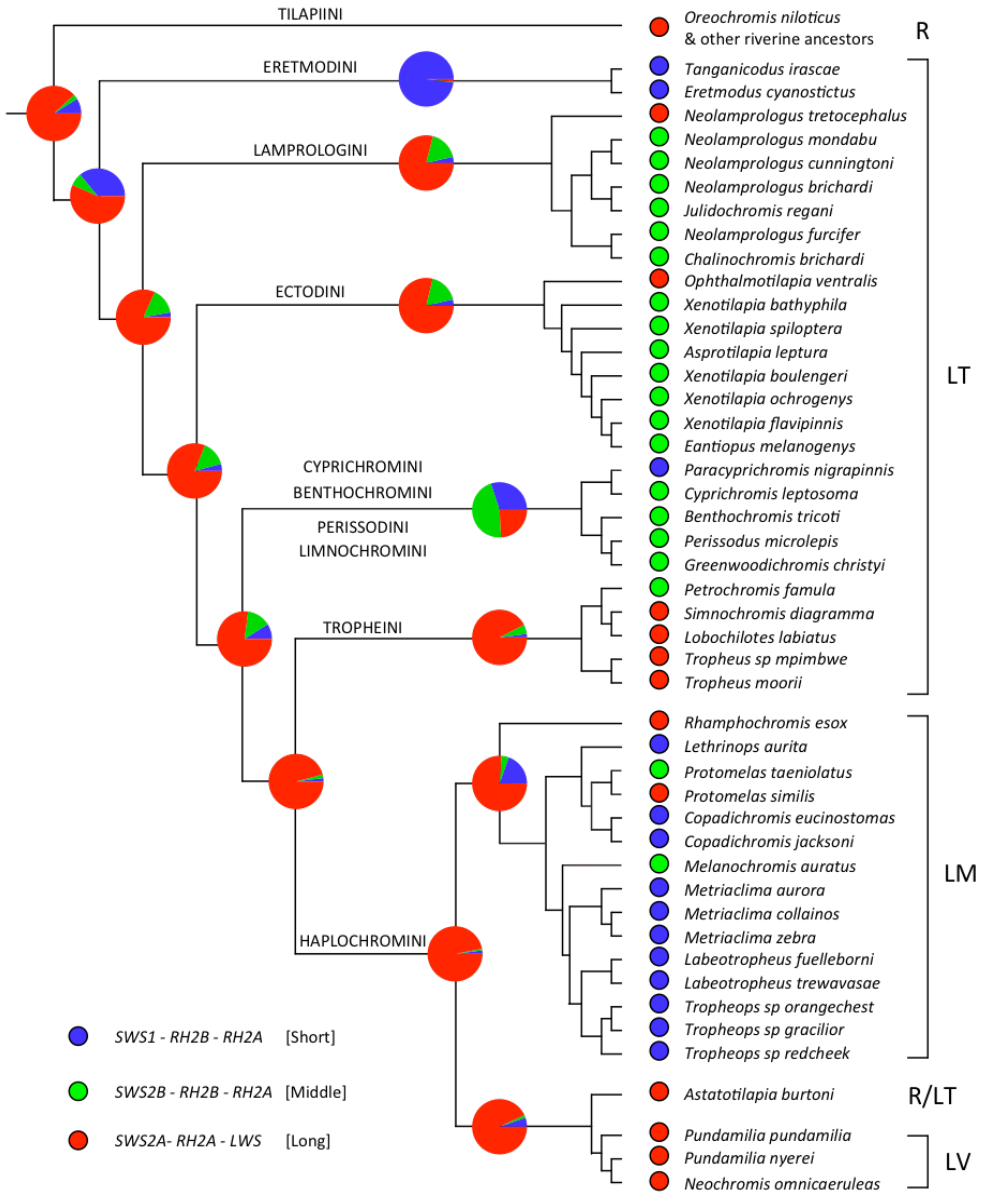
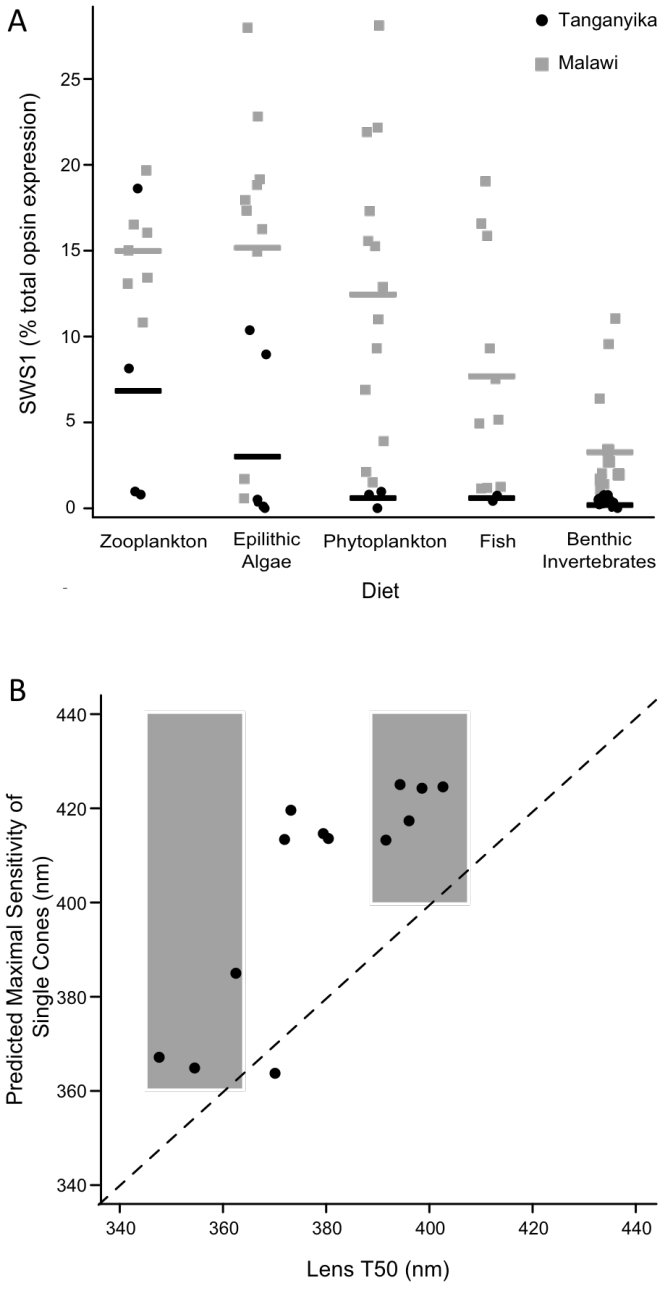


Figure 3-4. Comparative analysis of opsin gene expression with foraging preference and lens transmittance. (A) Mean *SWS1* (ultraviolet) opsin expression is higher among zooplanktivorous cichlid species than benthivorous ones (means indicated black bars). A similar pattern is observed among cichlids from LM (gray bars and boxes) [see Chapter 1]. (B) Regression of predicted maximal sensitivity (PS_{max}) of single-cone photoreceptors and lens transmittance (T50). Dotted line indicates $x = y$. The distribution of lens T50 and single-cone PS_{max} among LM cichlids is indicated with gray boxes (LM lens transmittance data from Hofmann et al. (2009)).



Chapter 4:

New Evidence for the Role of Heterochrony in the Repeated Evolution of Cichlid Opsin Expression

Kelly E. O'Quin, Adam R. Smith, Anit Sharma, and Karen L. Carleton

See Appendix 3 for all supplementary tables (Table S4-1) referenced in this chapter.

This chapter is published as:

O'Quin KE, Smith, AR, Sharma A, Carleton KL (2011) New evidence for the role of heterochrony in the repeated evolution of cichlid opsin expression. *Evolution & Development* 13: 193-203.

Abstract

Lake Malawi cichlids have undergone heterochronic shifts in the expression of their cone opsin genes, the genes responsible for color vision. These shifts have generated species with short-, middle-, and long-wavelength-sensitive cone photoreceptors and visual systems. However, it is unclear when during the evolution of African cichlids these shifts occurred, or whether they could account for similar short and middle wavelength-sensitive profiles among unrelated cichlids in Lake Tanganyika. To address these questions, we surveyed opsin expression in developing fry of two African cichlids, *Astatotilapia burtoni* from Lake Tanganyika and *Melanochromis auratus* from Lake Malawi. We found that *A. burtoni* expresses a series of three different single cone opsins over the course of development, while *M. auratus* exhibits variation in the expression of only two. Neither *A. burtoni* nor *M. auratus* exhibits much variation in the expression of its double cone opsins. These patterns reveal that *A. burtoni* exhibits progressive development in the sensitivity of its single cone photoreceptors, but direct development in the sensitivity of its double cone photoreceptors. *M. auratus* exhibits neotenic development in the sensitivity of both photoreceptor sets. Given the intermediate phylogenetic placement of *A. burtoni* between cichlids from Lakes Tanganyika and Malawi, our results suggest that the ancestor of Lake Malawi's cichlids exhibited a progressive developmental pattern of opsin expression. These results indicate that the heterochronic shifts which produced the short and middle wavelength-sensitive profiles of Lake Malawi's cichlids occurred recently, and suggest that the presence of similar profiles among Lake Tanganyika's cichlids are due to parallel heterochronic shifts.

Introduction

Heterochrony occurs during development when the appearance of one trait is changed relative to the appearance of another (McKinney and McNamara 1991; Gilbert 1997; West-Eberhard 2003). This process results in the "shifting of characters from one part of ontogeny to another" relative to the ancestral state (Valentine 1977). Two common forms of heterochrony include neoteny and direct development. Neoteny is the retention of larval or juvenile traits into the adult phase. Direct development is the opposite, the loss of larval or juvenile traits, resulting in the immediate appearance of the adult form of a trait early in development (Gilbert 1997). Examples of heterochrony abound in the literature, and include the evolution of morphological (Raff 1987) and behavioral traits (Garièpy et al. 2001), as well as gene expression (Wray and McClay 1989). Heterochrony is postulated to play an important role in phenotypic evolution, since it can generate new combinations of adult and larval phenotypes for selection to act upon (Gould 1977; McKinney and McNamara 1991; West-Eberhard 2003). This is especially true concerning the rapid or repeated evolution of new phenotypes, as is common in many adaptive radiations (West-Eberhard 2003).

African cichlids are a group of freshwater teleost fish that comprise the largest vertebrate adaptive radiation on earth (Sturmbauer 1998; Kocher 2004; Seehausen et al. 2008). These fishes are found throughout the rivers and lakes of Africa, but most species are concentrated within the three East African Great Lakes: Lakes Tanganyika (LT), Malawi (LM), and Victoria (LV). Cichlids from these lakes exhibit unparalleled diversity in many morphological, behavioral, and physiological

phenotypes; among others, these traits include impressive variation in the expression of seven cone opsin genes, the genes responsible for color vision (Carleton, Kocher 2001; Carleton et al. 2005; Parry et al. 2005; Hofmann et al. 2009; O'Quin et al. 2010). Variation in cichlid opsin expression has resulted in the evolution of many closely related cichlids that have photoreceptors sensitive to very different regions of the visible light spectrum (Parry et al. 2005; Jordan et al. 2006). Additionally, these patterns of opsin expression have evolved in parallel among unrelated cichlids from Lakes Tanganyika and Malawi (O'Quin et al. 2010) in response to important environmental and behavioral factors like foraging preference and the ambient light environment (Seehausen et al. 2008; Hofmann et al. 2009; O'Quin et al. 2010) [see Chapters 1 and 2]. This observation suggests that opsin expression diversity is the result of rapid, adaptive evolution in cichlids. A depiction of the broad phylogenetic relationships among cichlids from Lakes Tanganyika, Malawi, and Victoria is presented in Figure 4-1A.

Among the haplochromine cichlids of LM and LV, variation in adult opsin expression appears to be the result of heterochronic shifts from the developmental patterns of opsin expression observed in their distant ancestor, the Nile tilapia (*Oreochromis niloticus*) (Carleton et al. 2008). *O. niloticus* fry exhibit a progressive developmental pattern of opsin expression: larvae begin development by expressing a short-wavelength-sensitive set of opsins (*SWS1-RH2B-RH2A*), switch to a medium-wavelength-sensitive set as fry (*SWS2B-RH2B-RH2A*), and end development by expressing a long wavelength-sensitive set that is also observed in adults (*SWS2A-RH2A-LWS*) (Carleton et al. 2008). This progressive developmental pattern of opsin

expression results in photoreceptors that change their maximal sensitivity throughout ontogeny, from short-, to middle-, to long-wavelengths of light (Figure 4-1B). These same larval, juvenile, and adult opsin sets/photoreceptor sensitivities are observed among numerous adult cichlids from LM and LV, though following either neotenic or direct developmental patterns, not the progressive pattern observed in *O. niloticus* (Carleton et al. 2008) (Figure 4-1A and B). However, previous studies have only examined developmental variation among species that express either the short- or long-wavelength opsin sets as adults, not those with the middle-wavelength set. The patterns of developmental opsin expression so far observed in African cichlids are summarized in Figure 4-1B.

Unrelated adult cichlids from LT and LM exhibit very similar patterns of opsin expression that have evolved in parallel (O'Quin et al. 2010). O'Quin et al. (2010) found that the use of the short- (*SWS1-RH2B-RH2A*) and middle- (*SWS2B-RH2B-RH2A*) wavelength-sensitive opsin sets evolved in parallel among LT and LM cichlids from ancestors predicted to express the long-wavelength (*SWS2A-RH2A-LWS*) opsin set as adults (Figure 4-1A) [see also Figure 2-3 following Chapter 2 (O'Quin et al. 2010)]. This observation led the authors to hypothesize that the parallel evolution of the short- (*SWS1-RH2B-RH2A*) and middle- (*SWS2B-RH2B-RH2A*) wavelength opsin sets could be due to parallel heterochronic shifts from ancestors that each expressed the long-wavelength opsin set as adults, but the short- and middle-wavelength sets as fry. However, since no larvae or fry were available for any of the Tanganyikan species used by O'Quin et al. (2010), this hypothesis was left untested. Thus, the goal of our present study is to address the following question: what was the

likely developmental pattern of opsin expression found in the ancestor of LM cichlids? If the ancestor of LM cichlids exhibited the direct or neotenic developmental patterns observed in other LM cichlids, this result would suggest that the heterochronic shifts that gave rise to the evolution of the various adult opsin expression patterns found in LM cichlids occurred long ago, after the initial split of this lineage from *O. niloticus*. However, if this ancestor exhibited progressive development similar to that observed in *O. niloticus*, this result would suggest that the heterochronic shifts occurred more recently. This latter result would also suggest that the short- and middle- wavelength opsin sets of adult LT and LM cichlids could be due to similar and independent heterochronic shifts in opsin expression. To address these questions, we surveyed opsin expression in developing fry of the haplochromine cichlid *Astatotilapia burtoni*. *A. burtoni* is a basal member of the haplochromine cichlid lineage, the lineage to which all LM and LV cichlids belong. However, *A. burtoni* is found primarily in Lake Tanganyika and its surrounding rivers, and it is phylogenetically intermediate between cichlids from LT and LM (Salzburger et al. 2002; Koblmüller et al. 2008) (Figure 4-1A). Some studies have also placed *A. burtoni* in a polytomy with cichlids from LM and LV, and others even suggest that *A. burtoni* is ancestral to these species (Takahashi et al. 2001; Seehausen et al. 2003). Additionally, A. *burtoni* is considered to be similar to the hypothetical ancestor of all haplochromine cichlids, including those in LM and LV, since it is a generalized omnivore that is found in both rivers and lakes (Fryer and Iles 1972). Thus, given its intermediate phylogenetic position between cichlids from LT and LM, as well as its presumed similarity to the ancestor of all haplochromine cichlids, *A. burtoni* is an

excellent species with which to determine the probable ancestral developmental opsin profile of LM cichlids.

Fortunately, due to many years of research by Fernald and colleagues, *A. burtoni* is also a model system for the study of behavior (Grosenick et al. 2007), phenotypic plasticity (Renn et al. 2008), and vision (Fernald, Liebman 1980; Hagedorn, Fernald 1992). Hence, much is known about the visual system of *A. burtoni.* Previous work has established that: (*i*) during development, the retina of *A. burtoni* is dominated by cone photoreceptors (Hagedorn et al. 1998) but is structurally complete by ~ 7 days post fertilization (dpf) (Hagedorn and Fernald 1992); (*ii*) the photoreceptors of adult *A. burtoni* are maximally sensitive to 455, 523, and 562 nm wavelengths of light (Fernald and Liebman 1980), corresponding to the expression of the long wavelength opsin set (*SWS2A*/*B*-*RH2A*-*LWS*) (O'Quin et al. 2010); (*iii*) *A. burtoni* rhodopsin expression and photoreceptor cell growth exhibit diurnal and circadian rhythms (Korenbrot, Fernald 1989; Chiu, Mack, Fernald 1995); and (*iv*) the spectral sensitivity of *A. burtoni*'s photoreceptors are matched to the wavelengths of light present in the ponds and rivers surrounding LT (Fernald and Hirata 1977).

Despite the use of *A. burtoni* as a model system for the study of vision for over 30 years, it is still unknown whether *A. burtoni* exhibits ontogenetic variation in opsin expression. Therefore, we sampled developing *A. burtoni* fry to determine whether or not this species exhibits developmental variation in the expression of its opsins. We also sampled developing fry of the LM cichlid *Melanochromis auratus*. Adults of *M. auratus* express the middle-wavelength-sensitive opsin set (*SWS2B*-*RH2B*-*RH2A*), which was not surveyed in a previous developmental series of Lake

Malawi cichlids (Carleton et al. 2008). Therefore, it is unclear whether this species and those like it exhibit heterochronic developmental shifts in opsin expression as do other LM cichlids. We hypothesize that *A. burtoni* will exhibit progressive developmental variation in opsin expression like the riverine cichlid *O. niloticus*, and that *M. auratus* will exhibit neotenic developmental variation like other Lake Malawi cichlids. The results of this study will add important developmental information to the impressive list already available for *A. burtoni*, and they should also address what role heterochrony has played in the parallel evolution of adult opsin expression in African cichlids.

Materials and Methods

Sampling

We examined developmental variation in opsin expression among a lab-reared strain of *Astatotilapia burtoni* and a recently wild-caught strain of *Melanochromis auratus*. *A. burtoni* adults were provided by Daphne Soares (University of Maryland, College Park), while *M. auratus* adults were collected in 2008 from Lake Malawi National Park, Cape Maclear, Malawi (14°01'27.09"S, 34°49'27.03"E). We then mated each species in our lab to generate lab-reared broods for developmental sampling. Although we have recently reported that lab-rearing can slightly affect patterns of opsin expression relative to wild-caught individuals (Hofmann et al. 2009), this does not appear to prevent developmental changes in opsin expression from occurring (Carleton et al. 2008). For each species, we sampled fry starting at ~10 days post fertilization (dpf), since previous results indicate that the retina is fully

formed by this time (Hagedorn and Fernald 1992). We then continued sampling every 1 – 2 weeks until the broods were ~ 72 dpf, when the adult opsin set is nearly fixed (Carleton et al. 2008). This sampling scheme corresponds to the late larval to middle juvenile developmental stages (Fujimura and Okada 2007, 2008). We sampled fish at approximately the same time each day to avoid any variation due to diurnal changes in opsin expression (Korenbrot and Fernald 1989). For the *A. burtoni* broods, we collected six fish for each sampling period. However, for the first two sampling periods (14 and 28 dpf), we collected 9 fish; we then combined these fish into three replicates of three fish in order to generate enough RNA for expression analysis. For the *M. auratus* broods, we collected two fish for each sampling period, including the first two periods. At each sampling period, we euthanized the fish with tricaine methanesulfonate (MS-222, Argent, Redmond, WA), dissected both retinas, and stored them in RNA-later (Qiagen®, Valencia, CA) until the time of expression analysis.

Opsin Expression Analysis

We measured the relative expression of the opsin genes following our previously published protocols (Spady et al. 2006; Carleton et al. 2008; Hofmann et al. 2009; O'Quin et al. 2010). Briefly, we extracted RNA from the dissected retinas of each fish using QIAshredder® and RNeasy Mini® kits (Qiagen, Valencia, CA) following the manufacturers protocols. For each sample, we then reverse transcribed 0.5 μg total RNA to cDNA with Superscript III (Invitrogen, Carlsbad, CA). We quantified opsin expression via real time quantitative PCR (RT-qPCR) using species- and opsin-specific Taqman® primers and probes (Spady et al. 2006; Carleton et al.

2008; Hofmann et al. 2009; O'Quin et al. 2010). Finally, we normalized the expression of each opsin to either a construct of oligos encoding each opsin's primer and probe sequences (Spady et al. 2006), or else a dilution series with known concentrations of a single opsin's primer and probe sequence (O'Quin et al. 2010). As in our previous studies, we quantified the expression of the genetically and functionally similar *RH2Aα* and *RH2Aβ* opsins jointly. For each opsin, we recorded expression as percent of total opsin expression; thus, for each sample total opsin expression sums to 100%. Following the quantification of relative opsin expression, we used the results to predict the wavelength of maximum sensitivity for each species' single and double cone photoreceptors. Cichlid opsins are expressed within two distinct photoreceptor cell types, single cones and double cones (Fernald 1981; Carleton 2009). These predictions assume that the short-wavelength-sensitive opsins (*SWS1*, *SWS2B*, and *SWS2A*) are expressed in single cones, while the middle- and long-wavelength-sensitive opsins (*RH2B*, *RH2A*, and *LWS*) are expressed in double cones. These assumptions are supported by microspectrophotometry of single and double cone photoreceptors (Fernald 1981; Parry et al. 2005; Jordan et al. 2006; Spady et al. 2006; Carleton 2009), and have been confirmed using *in situs* (B Dalton, TW Cronin, KL Carleton, unpublished data). Therefore, following our previous studies (Carleton et al. 2008; Hofmann et al. 2009; O'Quin et al. 2010) we predict the maximal sensitivity of single and double cone photoreceptors using the formula:

$$PS_{max,\,C} = \frac{\Sigma(f_i \lambda_i)}{\Sigma f_i}$$

where f_i is the percent expression in of the i^{th} opsin in either single- or double-cones

(C), and λ_i is the wavelength of maximum absorbance (λ_{max}) of the i^{th} opsin in *O. niloticus* (Spady et al. 2006).

We perform two estimates of predicted sensitivity, one for the single-cones photoreceptors (based on *SWS1*, *SWS2B*, and *SWS2A* expression), and another double-cone photoreceptors (based on *RH2B*, *RH2A*, and *LWS* expression). There are two benefits to using these predictions. First, these estimates provide a useful summary statistic that describes multivariate changes in groups of opsins that exhibit evolutionarily and developmentally correlated patterns of expression (Carleton et al. 2008; Hofmann et al. 2009; O'Quin et al. 2010). Second, these predictions provide a more biologically relevant variable for analysis, since ultimately our goal is to determine how changes in opsin expression affect overall retinal sensitivity. Carleton et al. (2008) reported a very tight relationship between percent opsin expression and photoreceptor absorbance in *O. niloticus*, and Carleton (2009) summarizes this for several Lake Malawi cichlids as well.

Statistical Analyses

We used a combination of linear and nonlinear regression to determine whether significant changes in opsin expression/predicted photoreceptor sensitivity occur during *A. burtoni* and *M. auratus* development. We interpreted statistically significant change in developmental opsin expression as indicative of the progressive developmental pattern of *O. niloticus*, and insignificant or small change in opsin expression as indicative of the neotenic and direct developmental patterns of Lake Malawi cichlids. We performed all analyses using the lm() and nls() functions in the basic "stats" package of the R statistical computing software v2.10.1 (R Development

Core Team 2009). Depending on the distribution of the resulting data, we generally fit one of four possible regression models to the data: (*i*) a linear regression model of the form $y = \beta_0 + \beta_1(age) + \varepsilon$, where β_0 is the intercept, β_1 is the linear coefficient, and ε is the residual error; (*ii*) a polynomial or curvilinear regression model of the form $y = \beta_0 + \beta_1(age) + \beta_2(age^2) + \varepsilon$, where β_2 is the quadratic coefficient; (*iii*) a two-parameter exponential regression model of the form $y = \alpha * (dpf\hat{}\beta) + \varepsilon$, where α is the intercept and β is the rate at which y changes from its initial value between $0<age<\infty$; and (*iv*) a three-parameter exponential regression model of the form $y = \alpha + \beta * \exp(-\gamma * age)$, where α is the asymptote at age = ∞, β is the range of the response between $0<age<\infty$, and γ is the rate at which y changes from its initial value between $0<age<\infty$ (Ratkowsky 1990; Quinn and Keough 2002).

Results

Astatotilapia burtoni

We collected 30 *A. burtoni* individuals across five different developmental ages, ranging from ~ 14 – 70 days post fertilization (dpf). We observed considerable variation in the expression of most opsins during ontogenesis in *A. burtoni,* particularly among opsins expressed in the single cones (*SWS1*, *SWS2B*, and *SWS2A*) (Figure 4-2A). As in *O. niloticus* (Carleton et al. 2008) the short-wavelength-sensitive (SWS) opsins of *A. burtoni* exhibit a nonlinear and progressive pattern of expression over developmental time. In both species, *SWS1* (ultraviolet) opsin is the predominant SWS opsin expressed among larval fry <20 dpf. *SWS1* expression drops exponentially during this larval period, while in the same period *SWS2B* expression

increases. After 20 dpf, around the onset of the juvenile stage, *SWS2B* is the main opsin expressed in single cones. *SWS2B* expression peaks early, around 30 dpf, then also begins to drop, though more slowly than *SWS1* expression. As *SWS2B* expression drops, *SWS2A* expression slowly increases. By ~ 70 dpf, *SWS2A* seems to become the dominant SWS opsin expressed in single cones. Among opsins expressed in double cones (*RH2B*, *RH2A*, and *LWS*), both *RH2A* and *LWS* are expressed highly throughout development, though both exhibit slight curvilinear patterns of increased expression over time **(**Figure 4-2A). *RH2B* is not expressed at any time during ontogenesis, nor is it expressed in the adult. This pattern of double cone opsin expression differs from that observed in *O. niloticus*, where *RH2B* is the dominant opsin expressed in larval fish (those <20 dpf), after which *LWS* expression rapidly climbs to high expression during the juvenile stage and remains on throughout the rest of development and adulthood. Also dissimilar is the slight increase in *RH2A* expression observed for *A. burtoni*, whereas in *O. niloticus RH2A* expression begins at 20% relative expression and then decreases slowly until adulthood (Carleton et al. 2008).

We note that Figure 4-2 illustrates changes in opsin expression with lines fitted using either linear or nonlinear regression, depending on the distribution of the data. These curves are only meant to highlight the trends in opsin expression and are not used for hypothesis testing, which we perform on predicted photoreceptor sensitivities below. We note also that although our analysis of development ends at ~70 dpf (middle juvenile stage), for comparison we include final relative expression values for adult *A. burtoni* from O'Quin et al. (2010). In general, the final values at ~

70 dpf closely match those in the adult, but in a few cases suggest that opsin expression may continue to increase (for example, *SWS2A*) or decrease (for example, *RH2A* and *LWS*). Despite this limitation, it is clear that, between 10 and 70 dpf, the opsins of *A. burtoni* exhibit considerable change in expression (Figure 4-2A).

Melanochromis auratus

We collected 22 *M. auratus* individuals between 10 and 72 dpf. In contrast to *O. niloticus* and *A. burtoni*, we observed considerable ontogenetic change in opsin expression for only three opsins, *SWS1*, *SWS2B*, and *RH2B*; the remaining opsins exhibited slight linear increases in expression, or no apparent increase at all (Figure 4-2B). Among opsins expressed in single cone photoreceptors, *SWS1* was the predominant opsin expressed until ~ 14 dpf; however, like *A. burtoni*, *SWS1* expression dropped exponentially during this period. At the same time, *SWS2B* expression rapidly increased. After ~ 14 dpf, *SWSB* was the SWS opsin with the highest expression in single cones. *SWS2B* was then the predominant SWS opsin expressed in *M. auratus* throughout the juvenile period. *SWS2A* was never expressed in developing or adult *M. auratus* (Hofmann et al. 2009; Hofmann et al. 2010b). For opsins expressed in the double cones, *RH2A* expression was the highest of all opsins throughout development, and even increased slightly into adulthood (Figure 4-2B). In contrast, *RH2B* expression was low prior to ~12 dpf but increased exponentially to 20% and remained at this level for the rest of development. Finally, this species also expressed the *LWS* opsin, which remained constant throughout development. The smaller shifts in opsin expression we observe for *M. auratus* are consistent with the patterns of developmental change seen in other haplochromine cichlids from lakes

Malawi and Victoria (Carleton et al. 2008).

Predicted Photoreceptor Sensitivity

The predicted single cone sensitivities of both *A. burtoni* and *M. auratus* varied nonlinearly with age in developing fry. In *A. burtoni*, singe cone PS_{max} increased from approximately 380 to 425 nm over the course of development (Figure 4-3A). We found that a two-parameter exponential regression model fit the distribution of observed values well; the estimated coefficients for this model and their associated significance values are listed in Table 4-1. Importantly, the parameter β, which here describes the rate of change in single cone PS_{max} between dpf = 0 and dpf = ∞, is significantly different from zero ($\beta = 0.068$, $t = 5.664$, $p < 0.001$; see Table 4-1). This result indicates that significant change in single cone PS_{max} occurs over *A. burtoni* development. However, we note also that variation in the single cone PS_{max} for *A. burtoni* was generally large, and we excluded one outlier at ~ 60 dpf from this analysis (this outlier is still shown in Figure 4A). For *M. auratus*, single cone PS_{max} increased exponentially from 390 to 418 nm between 10 and 20 dpf (larval development), but then remained at 418 nm for the remainder of development (juvenile period) and into adulthood. We fitted a three-parameter exponential curve to these data and once again found that the rate of change parameter was significantly different from zero ($\gamma = 0.157$, $t = 5.661$, $p < 0.001$). We excluded one outlier at ~ 40 dpf from this analysis. In Figure 4-3 we include the single cone PS_{max} of *O. niloticus* (Carleton et al. 2008) for comparison with *A. burtoni* and *M. auratus*. The results show that in *O. niloticus*, single cone PS_{max} increases steadily between 10 and 70 dpf, paralleling closely the increase observed in *A. burtoni*. The single cone PS_{max} of *M.*

auratus, however, levels out quickly relative to these other two species. Thus, *A. burtoni* exhibits progressive development of single cone sensitivity, while *M. auratus* exhibits progressive development up to 20 dpf, but neotenic development thereafter (Figure 4-3A).

The double cone PS_{max} of *A. burtoni* and *M. auratus* exhibited much less variation than the PS_{max} of single cone photoreceptors (Figure 4-3B). For *A. burtoni*, the PS_{max} of double cones ranged from 544 to 550 nm, and decreased slightly over time. However, despite the small magnitude of this change, we still observed a significant curvilinear relationship between double cone PS_{max} and age in *A. burtoni* (Figure 4-3B; Table 4-1). The double cone PS_{max} of *M. auratus* also varied little; values ranged from 511 to 523 nm. In contrast to *A. burtoni*, however, we found no significant relationship between double cone PS_{max} and age in *M. auratus*, linear or otherwise ($\beta_1 = -0.019$, $t = 0.628$, $p = 0.538$; Table 4-1). We excluded one observation at ~ 40 dpf as an outlier from the analysis of *M. auratus* double cone PS_{max}. Comparison of *A. burtoni* and *M. auratus* double cone PS_{max} to those of *O. niloticus* clearly illustrates neoteny in the double cone sensitivity of *M. auratus* and direct development in *A. burtoni* (Figure 4-3B).

Finally, we note that all models produced errors that were approximately normally distributed with equal variances. We used the Bonferroni-corrected significance threshold for 10 comparisons ($\alpha = 0.05/10 = 0.005$) when determining the significance of the various coefficients in our models (see Table 4-1). All opsin expression results and predicted photoreceptor sensitivities reported for *A. burtoni*, *M.*

auratus, and *O. niloticus* are available as supplemental information in Appendix 3 (Table S4-1 [Appendix 3]).

Discussion

Developmental Variation in Cichlid Opsin Expression and Photoreceptor Sensitivity

We found that the Lake Tanganyika (LT) haplochromine cichlid *Astatotilapia burtoni* exhibits significant nonlinear variation in the expression of several cone opsin genes during the course of its development (Figure 4-2A). *A. burtoni* exhibits progressive development of single cone opsin expression (*SWS1*, *SWS2B*, and *SWS2A*) in a manner paralleling that seen in *O. niloticus* (Carleton et al. 2008). However, the expression of double cone opsins in developing *A. burtoni* fry largely follow a pattern of direct development, in contrast to the progressive developmental pattern seen in *O. niloticus*. This difference is due to a lack of *RH2B* expression at any developmental stage in *A. burtoni*. Thus, it is possible that *RH2B* has become a pseudogene in this species. O'Quin et al. (2010) generated species- and opsin-specific primers and probes for each opsin, including *RH2B*, so it is unlikely that this result is due to error in the quantification of opsin expression. Also, little change is seen in the expression of the other two double cone opsins, *RH2A* and *LWS*, suggesting that the *RH2B* opsin truly is not expressed at any age in this species (Figure 4-2A). Together, the observed variation in single and double opsin expression is expected to produce single and double cone sensitivities that also vary significantly with developmental age (Figure 4-3A), although this variation is much greater for single cone sensitivities. The resulting estimates of maximal

photoreceptor sensitivity suggest that the maximal sensitivity of single cone photoreceptors in *A. burtoni* changes in a progressive manner, similar to that observed in *O. niloticus* (Figure 4-3). In contrast, the wavelengths of light that double cones are maximally sensitive to do not appear to change in this species.

We observed similar patterns of developmental variation for the Lake Malawi (LM) haplochromine *Melanochromis auratus*, but with much smaller changes. We primarily observed significant nonlinear change in the expression of single cone opsins for *M. auratus* (Figure 4-2B). *M. auratus* exhibits progressive developmental expression of the two shortest-wavelength-sensitive SWS opsins, *SWS1* and *SWS2B*. The *SWS2A* opsin is never expressed in this species as it is in *A. burtoni* and *O. niloticus*. Additionally, the observed changes occur quickly, settling on *SWS2B* expression by ~14 dpf, prior to the onset of the juvenile period. In contrast to single cone opsins, we found that double cone opsin expression exhibits little change in this species, resulting in predicted double cone sensitivities that remain static across development. These patterns produce predicted single cone sensitivities that change progressively early in development, but stop short of the full progressive pattern seen in *A. burtoni* and *O. niloticus*. This pattern is similar to, though a little more pronounced than, that observed for other neotenic LM species (Carleton et al. 2008). Predicted maximal double cone sensitivities do not change in this species and are clearly neotenic with respect to the pattern observed in *O. niloticus* (Figure 4-3B).

The single and double cone photoreceptor cells of cichlids are arranged in a highly ordered mosaic within the retina (Fernald and Liebman 1980). In *A. burtoni*, four pairs of long and medium wavelength-sensitive double cones surround one short

wavelength-sensitive single cone; this pattern is then repeated throughout the retina. Fernald (Fernald 1981) reported that the double cones of *A. burtoni* exhibit an alternating symmetry, with one of the double cones maximally sensitive to middle-wavelength light and the other to long-wavelength light. This pattern suggests that the green and red sensitive opsins (*RH2A* and *LWS*) should be expressed equally in the retina of *A. burtoni*, but our results indicate that this is not the case. We find that *LWS* expression is always greater than *RH2A* expression across all developmental stages in *A. burtoni*. This observation suggests that not all pairs of double cones demonstrate alternating symmetry in *A. burtoni*, and that some might express either *LWS* or *RH2A* exclusively. An excess of *LWS* expression would suggest identical *LWS*/*LWS* cone pairs in addition to the alternating *LWS*/*RH2A* pairs. Indeed, *LWS*/*LWS* double cone pairs have been observed in cichlids from Lake Victoria (Carleton et al. 2005) and *O. niloticus* (Carleton et al. 2008)B Daleton, TW Cronin, KL Carleton, unpublished data). Additionally, the number of identical and alternating double cone cells varies across the retina in several fish species, including cichlids (Levine et al. 1979). This variation is likely related to differences in the spectral distribution of up- and down-welling light, as well as the spectral requirements of specific visual tasks. The analysis we report here summed opsin expression across the entire retina, and therefore cannot determine what patterns of double cone opsin expression are present in *A. burtoni*. In the future we plan to use *in situ* hybridization to examine topographical variation in opsin expression across the retina of developing and adult cichlids.

Finally, we note that our use of opsin expression to predict photoreceptor

sensitivity reflects our assumption that the changes in opsin expression we observe are translated into similar changes at the cone level—that is to say, opsin expression fluctuates qualitatively within the photoreceptors of developing cichlids. Several previous studies have examined retinal morphology in developing cichlids (van der Meer, Anker 1986; Hagedorn, Fernald 1992; van der Meer 1995; Braekevelt, Smith, Smith 1998; Hagedorn et al. 1998) and found no large gain or loss of cone cells during development or growth. Additionally, cichlids do not possess accessory corner cones for the expression of SWS opsins, as do many salmonid fishes (Fernald, Liebman 1980; Fernald 1981; Allison et al. 2003). These observations suggest that the changes in opsin expression we observe must occur within the photoreceptors already present in the retina. The results of Carleton et al. (2008), who measured both opsin expression and photoreceptor sensitivity in developing *O. niloticus* fry, confirm this hypothesis. These authors found that developmental changes in opsin gene expression were accompanied by corresponding changes in photoreceptor sensitivity. Thus, we believe the patterns of developmental photoreceptor sensitivity we predict for *A. burtoni* and *M. auratus* are likely to be reflected in any physiological measurements. Future work may confirm this assumption in a larger panel of African cichlid species, including those from Lake Tanganyika.

The Evolution of Adult Opsin Expression Diversity through Heterochrony

A. burtoni, *M. auratus*, and other haplochromine cichlids clearly exhibit heterochronic shifts in opsin expression and predicted photoreceptor sensitivity relative to *O. niloticus*. But when did these shifts occur? Our analysis of developmental variation in opsin expression for the LT haplochromine *Astatotilapia*

burtoni reveals that this species also exhibits a progressive developmental pattern of single cone opsin expression. This result supports the hypothesis that the heterochronic shifts observed in LM and LV cichlids occurred *after* their split from other basal haplochromines. This pattern of evolution is summarized in Figure 4-1C. Since both *O. niloticus* and *A. burtoni* exhibit progressive development of single cone sensitivity and express the long-wavelength opsin set as adults, we infer that other phylogenetically intermediate species that are also predicted to express the long wavelength opsin set [e.g., those ancestral nodes between *O. niloticus* and *A. burtoni* in Figure 4-1A) [see also Figure 2-3 following Chapter 2] did so following a similar developmental progression. It is unclear if this is also the case for double cone sensitivities, although we believe it is. Since haplochromine cichlids from LM exhibit functional *RH2B* expression, loss of expression of this gene is probably specific to the LV/*A. burtoni* lineage, and therefore is not indicative of the ancestral haplochromine.

Taken together, our results suggest that the heterochronic shifts in developmental opsin expression observed in LM and LV cichlids likely occurred within the last 5 MY, after the split of these species from the basal haplochromine cichlid lineage (Koblmüller et al. 2008). Therefore, the presence of similar short- and middle-wavelength-sensitive opsin sets among adult LM and LT cichlids could be due to parallel heterochronic shifts in developmental opsin expression. However, developmental sampling of LT cichlids with the short-, middle-, and long-wavelength opsin sets will be necessary to confirm this hypothesis. In particular, the LT lineage Tropheini is sister to the haplochromine lineage (Salzburger et al. 2002), and the

majority of these species express the long wavelength opsin set as adults (O'Quin et al. 2010). Our results suggest that these species should exhibit progressive development of single and double cone photoreceptor sensitivities. If this hypothesis is correct, then the inference that most ancestral species also exhibited progressive development will be bolstered. Similarly, sampling of the basal haplochromine *Astatotilapia calliptera*, which is found along the marshy shores of LM, could also bolster our conclusion. Such analyses would provide additional evidence that the parallel evolution of adult opsin expression among LM and LT cichlids is the result of parallel heterochronic shifts in opsin expression.

Foraging, Vision, and Heterochrony in African Cichlids

In addition to changes in opsin expression, many cichlids and other teleosts undergo changes in foraging preference over the course of development (Fryer 1959; Fryer, Iles 1972; Wanink, Joordens 2007; Zengeye, Marshall 2007). In many teleost fishes, fry typically feed on zooplankton and phytoplankton, juveniles feed on algae and macroinvertebrates, and adults finally settle on larger invertebrates and other fish or their eggs (Fryer and Iles 1972). Fryer (Fryer 1959) even speculated that cichlids that forage in shoals for zooplankton exhibit paedomorphic or neotenic foraging behaviors. Although we do not have specific information on the larval and juvenile diets of *A. burtoni*, *M. auratus*, or *O. niloticus*, we do know that these species differ in their predominant diet as adults. Adult *M. auratus* primarily feed on algae and plankton while adult *A. burtoni* and *O. niloticus* are both omnivorous, but frequently feed on fish or benthic invertebrates (Brichard 1978; Ribbink et al. 1983).

In African cichlids, divergent opsin expression profiles and photoreceptor

sensitivities represent an adaptation to divergent foraging preferences (Munz, McFarland 1977; Jordan et al. 2004; Hofmann et al. 2009; O'Quin et al. 2010) [see Chapters 1 and 2]. In both LM and LT cichlids, *SWS1* (ultraviolet [UV]) opsin expression and UV single cone photoreceptor sensitivity is significantly associated with planktivory (Jordan et al. 2004; Hofmann et al. 2009; O'Quin et al. 2010). UV sensitivity presumably helps these and other species detect UV-absorbing zooplankton against the bright down-welling light (Novales-Flamarique and Hawryshyn 1994). Here, we show that ultraviolet sensitivity is likely to be highest in the earliest, or larval, stages of development in *A. burtoni* and *M. auratus*, a pattern that is also observed in *O. niloticus* (Figures 3-2 and 3-3) (Carleton et al. 2008). In contrast, LM and LT cichlids that forage on fish and benthic invertebrates express very little *SWS1* opsin, and typically exhibit the long wavelength opsin set (*SWS2A-RH2A-LWS*) (Munz, McFarland 1977; Carleton et al. 2008; Hofmann et al. 2009; O'Quin et al. 2010). This is the same opsin set found in adult *O. niloticus* and *A. burtoni*, both which feed on fish and benthic invertebrates (Carleton et al. 2008). Thus, not only is there an evolutionary correlation between adult opsin expression and foraging preference in cichlids, but there is a developmental correlation between these two traits as well. We postulate that heterochronic shifts in both age-specific patterns of opsin expression and trophic preference are responsible for maintaining the correlated evolution of these two traits in cichlid adults. Heterochrony may therefore provide a developmental basis for evolutionary change in many trophic specializations in cichlids, while maintaining necessary foraging-specific adaptations of the visual system. It is even possible that the correlated evolution of these two

traits in adults is due to a single heterochronic mechanism that simultaneously affects both traits.

Conclusions

Heterochrony is evolutionary change in the developmental timing of expression for a trait (Valentine 1977; West-Eberhard 2003), which is postulated to play an important role in phenotypic evolution. Here we further investigated evidence for heterochrony in the visual system of African cichlid fishes, one of the most phenotypically diverse families of fish on earth (Kocher 2004). We found that the haplochromine cichlid *Astatotilapia burtoni* undergoes considerable developmental change in the expression of several opsins associated with single-cone or short-wavelength photoreceptor sensitivity. Additionally, the reduced variation found in the Lake Malawi cichlid *Melanochromis auratus* confirms that LM cichlids generally exhibit heterochronic developmental shifts in opsin expression and photoreceptor sensitivity (Figure 4-3). Given the relatively basal position of *A. burtoni* to the African cichlids of Lakes Malawi and Victoria, our results suggest that the heterochronic shifts which gave rise to the short- and middle-wavelength-sensitive opsin sets in these latter species occurred recently, after the split of these lineages from *A. burtoni* (Figure 4-1). Therefore, the presence of similar short- and middle-wavelength-sensitive opsin sets among African cichlids in Lake Tanganyika may be due to parallel heterochronic shifts in these species (O'Quin et al. 2010), although further work is needed to conclusively demonstrate this. Our results provide many new insights into the visual system of *A. burtoni,* suggest an important role for

heterochrony in the evolution of divergent foraging and visual adaptations in cichlids, and raise many new questions and hypotheses for future study.

Tables

Table 4-1. Statistical significance of linear and nonlinear coefficients estimated for *A. burtoni* and *M. auratus* predicted photoreceptor sensitivities.

Species	Photoreceptor	Model	Coefficient	Estimate	df	t-value	p-value
A. burtoni	Single Cone	$\alpha * (dpf^{\wedge}\beta)$	α	322.702	21	21.879	< 0.001
			β	0.068	21	5.664	< 0.001
	Double Cone	$\beta_0 + \beta_1(dpf) + \beta_2(dpf^2)$	β_0	553.400	20	517.151	< 0.001
			β_1	-0.318	20	5.962	< 0.001
			β_2	0.003	20	4.747	0.001
M. auratus	Single Cone	$\alpha - \beta * \exp(-\gamma * dpf)$	α	419.141	18	526.573	< 0.001
			β	126.781	18	3.349	0.004
			γ	0.157	18	5.661	< 0.001
	Double Cone	$\beta_0 + \beta_1(dpf)$	β_0	516.056	19	387.899	< 0.001
			β_1	-0.019	19	0.628	0.538

Figures

Figure 4-1. Evolution of opsin expression in African cichlids. (A) Phylogeny representing the broad relationship of African cichlids in Lakes Tanganyika (LT), Malawi (LM), and Victoria (LV), with special emphasis on *Oreochromis niloticus* and *Astatotilapia burtoni*. Colored dots represent the adult opsin expression profiles found among cichlids in each lake, or else inferred for ancestral lineages [see also Figure 2.3]. These profiles are short- (blue-green), middle- (green), and long- (red) wavelength sensitive (WS). We denote with an asterisk (*) the node that represents the universal ancestor of all haplocomine cichlids. (B) Prior to our current study, these are the known patterns of developmental photoreceptor sensitivity for LM cichlids and *O. niloticus* (Carleton et al. 2008). Ages shown are larval (L), juvenile (J), and adult (A). The purpose of the present study is to determine the developmental patterns of photoreceptor sensitive for those groups marked by a question mark (?). (C) After our current study, these are the inferred patterns of developmental photoreceptor sensitivity in African cichlids from LT, LM, and LV.

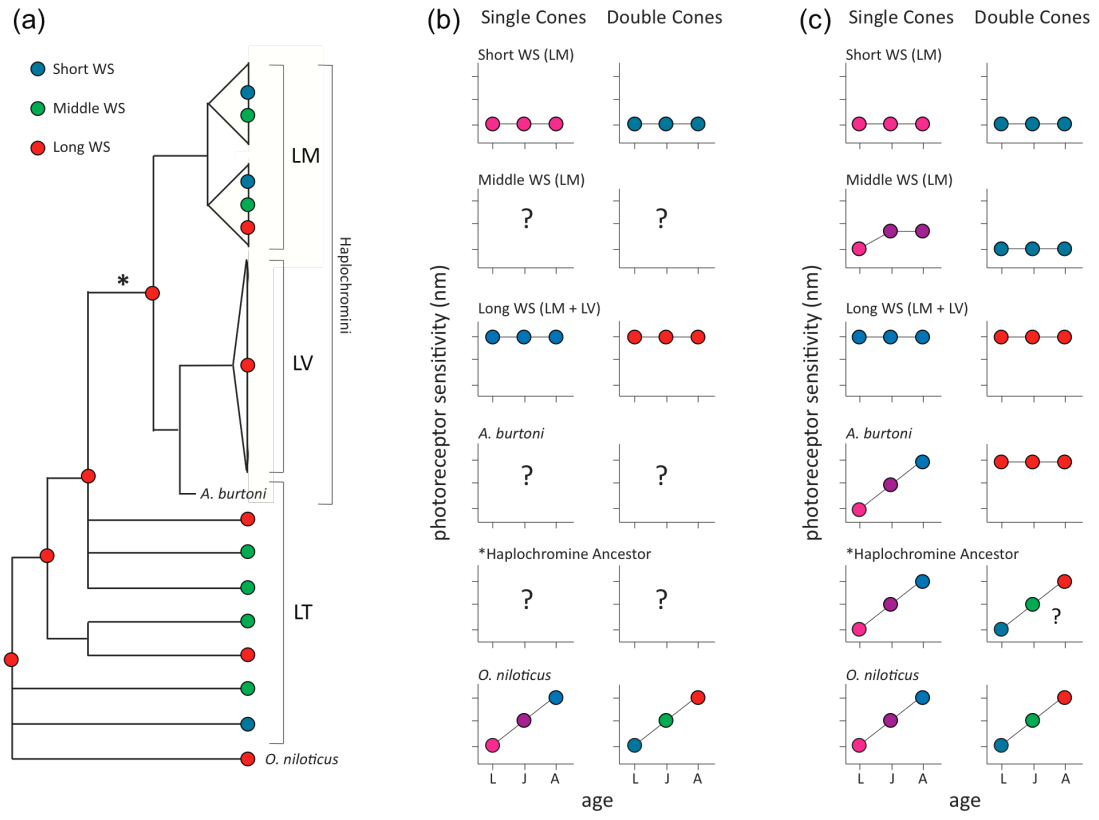


Figure 4-2. Ontogenetic variation in opsin gene expression for two African cichlids, (A) *Astatotilapia burtoni* and (B) *Melanochromis auratus*. Fish were sampled between approximately 10 and 70 dpf. Onset of the juvenile stage is denoted with a vertical dashed line; onset of the adult stage is denoted with a vertical solid line. *A. burtoni* adult opsin expression profile from O'Quin et al. (2010), *M. auratus* from Hofmann et al. (2009) [see Chapters 1 and 2]. The wavelength of maximum absorbance (λ_{max}) for each opsin is shown beneath its name; each λ_{max} value was determined previously based on the opsin sequences of *O. niloticus* (Spady et al. 2006).

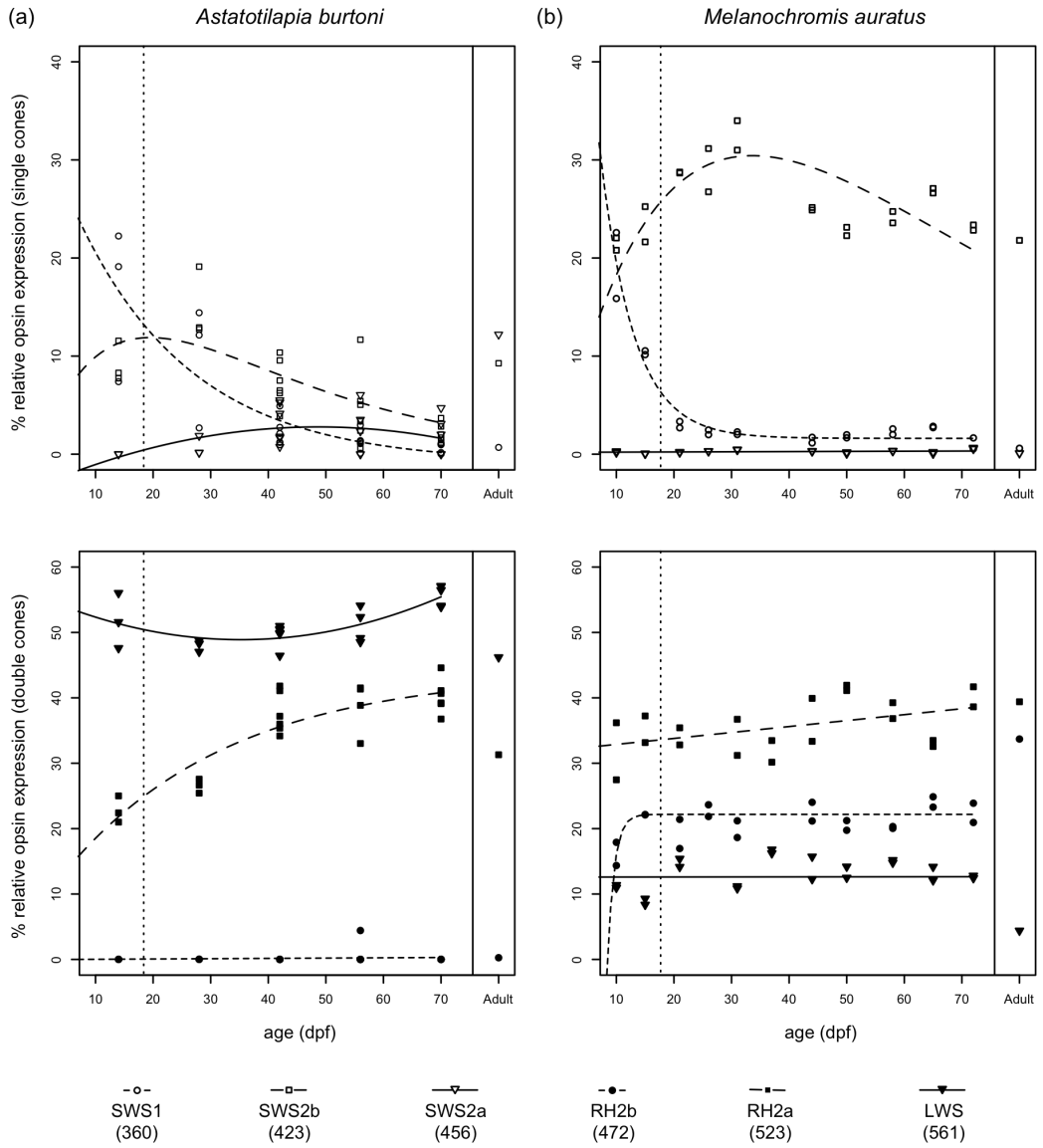
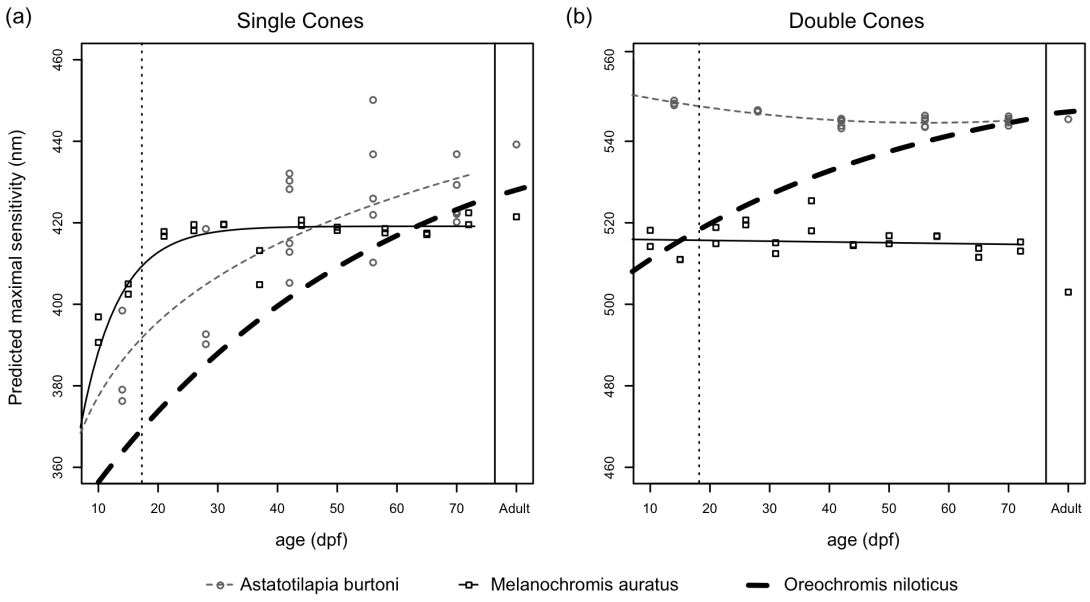


Figure 4-3. Ontogenetic variation in the predicted maximal sensitivity of single- and double-cone photoreceptors for (A) *Astatotilapia burtoni* and (B) *Melanochromis auratus*. The sensitivity of single- and double-cone photoreceptors observed for *O. niloticus* during the same developmental period is indicated with a thick dashed line (Carleton et al. 2008). The predicted maximal sensitivity of *A. burtoni* single-cone photoreceptors changes dramatically over the course of development and largely parallels the progressive developmental pattern observed in *O. niloticus*. The predicted maximal sensitivity of *M. auratus* single cones changes less, suggesting a slightly neotenic developmental pattern. For both *A. burtoni* and *M. auratus*, the predicted maximal sensitivity of their double-cone photoreceptors changes very little, indicating direct and neotenic development, respectively. Linear and nonlinear formulas used to construct regression lines are presented in Table 4.1.



Chapter 5:

Divergence in *cis*-regulatory sequences surrounding the opsin gene arrays of African cichlid fishes

Kelly E. O'Quin, Daniel Smith, Zan Naseer, Jane Schulte, Samuel D. Engel, Yong-Hwee E. Loh, J. Todd Streelman, Jeffery L. Boore, and Karen L. Carleton

See Appendix 4 for all supplementary tables and figures (Supplementary Tables S5-1 — S5-6 and Figures S5-1 — S5-5) referenced in this chapter.

This chapter has been revised and is accepted for publication as:

O'Quin KE, Smith D, Naseer Z, Schulte J, Engel SD, Loh Y-HE, Streelman JT, Boore JL, Carleton KL (2011) Divergence in *cis*-regulatory sequences surrounding the opsin gene arrays of African cichlid fishes. *BMC Evolutionary Biology*.

Abstract

Divergence within *cis*-regulatory sequences may contribute to the adaptive evolution of gene expression, but functional alleles in these regions are difficult to identify without abundant genomic resources. Among African cichlid fishes, the differential expression of seven opsin genes has produced adaptive differences in visual sensitivity. The loci that control opsin expression in cichlids are unknown, but quantitative genetic analysis suggests that *cis*-regulatory alleles may contribute to this variation. Here, we sequence BACs containing the opsin genes of two African cichlids, *Oreochromis niloticus* and *Metriaclima zebra*. We use phylogenetic footprinting and shadowing to examine divergence in conserved non-coding elements, promoter sequences, and 3'-UTRs surrounding each opsin in search of *cis*-regulatory sequences that may influence cichlid opsin expression. We identified 23 conserved non-coding elements surrounding the opsins of cichlids and other teleosts, including two homologous to a known opsin enhancer and a retinal microRNA. Most conserved elements contained computationally-predicted binding sites that correspond to transcription factors that function in vertebrate opsin expression, but *O. niloticus* and *M. zebra* were significantly divergent in only four of these elements. Similarly, we found a large number of relevant transcription factor binding sites within each opsins' proximal promoter, and identified five opsins that were considerably divergent in both expression and the number of transcription factor binding sites shared between our two focal species. We also found several conserved and non-conserved microRNA target sites within the 3'-UTR of each opsin in cichlids, including two that differ significantly between *O. niloticus* and *M. zebra*.

Finally, we examined interspecific divergence in the proximal promoters of five opsins in 18 phenotypically diverse cichlids from Lake Malawi. We found that the promoters examined were highly conserved with some evidence of CRX transcription factor binding site turnover, and we found three SNPs with weak association to cichlid opsin expression. This study is the first to systematically search the opsins of cichlids for putative *cis*-regulatory sequences. We found that many putative regulatory regions are highly conserved across a large number of phenotypically diverse species, but we did identify 9 divergent sequences that stand out as candidates for future functional analysis in cichlids.

Introduction

Adaptive phenotypic evolution may result either from protein-coding mutations that modify the structure and function of genes, or from regulatory mutations that alter the timing, location, or amount that genes are expressed (Carroll 2005; Hoekstra and Coyne 2007; Wray 2007). Although examples of protein-coding mutations that contribute to phenotypic evolution are well known (Jessen et al. 1991; Yokoyama et al. 1999; Hoekstra et al. 2006), examples of regulatory mutations that also affect phenotypic adaption are less well-known, but no less important (Tishkoff et al. 2007; Jeong et al. 2008; Chan et al. 2010). One class of these mutations, *cis*-regulatory mutations, are found in close proximity to the genes they regulate and function by altering the binding of transcription factors necessary for gene expression. *Cis*-regulatory mutations exhibit several features that make them ideally suited for adaptive phenotypic evolution, including codominance (Lemos et al. 2008) and

modularity (Jeong et al. 2008). These features make *cis*-regulatory mutations efficient targets for natural selection (Hartl and Clark 2006) and limit the negative consequences of pleiotropy that presumably affect *trans*-regulatory and protein-coding mutations. Finally, since *cis*-regulatory mutations may underlie many of the adaptive and disease phenotypes found in nature, identifying these alleles remains an important goal of evolutionary genetics. However, indentifying *cis*-regulatory mutations can be challenging without abundant functional genomic resources, since the transcription factor binding sites (TFBS) they affect are small, lack strict conservation, and are found in under-annotated regions of the genome (Hoekstra and Coyne 2007; Wray 2007).

The location of *cis*-regulatory sequences can be near-to or far-from the genes they regulate. Promoter sequences found directly upstream of genes can harbor *cis*-regulatory alleles (Yuh et al. 1998; Berman et al. 2002), as can enhancer or repressor elements located many kilobases away (Tuan et al. 1989; Ebert et al. 1995). *Cis*-regulatory sequences can even reside within the untranslated regions (UTRs) of genes, where they alter the binding of microRNAs that regulate gene expression following transcription (Kloc et al. 2000; Chen and Rajewsky 2006). But where ever their location, one method commonly used to identify *cis*-regulatory sequences is phylogenetic footprinting (Gumucio et al. 1996). Phylogenetic footprinting compares DNA surrounding some gene(s) of interest among numerous divergent taxa in hopes of identifying non-coding regions that are highly conserved. By the very nature of their conservation, these conserved non- coding elements (CNEs) stand out as candidate regulatory sequences, since conservation is often used to indicate function.

Once candidate regulatory sequences have been identified, the method used to identify putative *cis*-regulatory alleles within them is differential phylogenetic footprinting, or phylogenetic shadowing (Gumucio et al. 1996; Boffelli et al. 2003). Phylogenetic shadowing compares putative regulatory sequences among closely related taxa in hopes of identifying sequence polymorphisms correlated with divergent expression of the target gene(s). Following their application, functional genomic analyses are necessary to validate the function of any candidate sequences or alleles identified by the phylogenetic footprinting and shadowing methods; but even by themselves, both methods can provide valuable insights into the location of potential *cis*-regulatory sequences and the transcription factors that bind them.

The goal of this study is to identify candidate *cis*-regulatory sequences that control opsin gene expression in African cichlid fishes. Opsins are a group of G protein-coupled receptors that confer sensitivity to light and mediate color vision (Wald 1968). African cichlids comprise a diverse clade of freshwater, teleost fish found throughout the lakes and rivers of Africa, including the three African Great Lakes, Lakes Tanganyika, Malawi, and Victoria (Kocher 2004; Seehausen 2006). Cichlids from Lakes Tanganyika and Malawi exhibit dramatic variation in their sensitivity to colored light (Carleton et al. 2006; Jordan et al. 2006; Carleton 2009). Species from these lakes exhibit retinal sensitivities that are maximally sensitive to short, middle, or long-wavelength spectra; in some cases, closely related species can differ in their maximal retinal sensitivity by over 100 nm (Jordan et al. 2006; Hofmann et al. 2009; O'Quin et al. 2010). This striking variation makes the cichlid visual system the most diverse vertebrate visual system on earth. Most variation in

cichlid color sensitivity is due to changes in the regulation of their cone opsin genes (Hofmann et al. 2009; O'Quin et al. 2010). Cichlids have seven cone opsin genes used for color vision; these opsins are *SWS1* (ultraviolet-sensitive), *SWS2B* (violet-sensitive), *SWS2A* (blue-sensitive), *RH2B* (blue-green-sensitive), *RH2A*α and *RH2A*β (green-sensitive), and *LWS* (red-sensitive) (Spady et al. 2006). Among different cichlid species, these opsins are alternatively co-expressed in three predominant groups, or palettes, to produce the three common visual pigment sets; these palettes are *SWS1-RH2B-RH2A* (short wavelength-sensitive), *SWS2B-RH2B- RH2A* (middle wavelength-sensitive), and *SWS2A-RH2A-LWS* (long wavelength- sensitive) (Hofmann et al. 2009). Cichlids exhibit several correlations between the expression of their opsins and important ecological variables, including foraging preference and ambient light intensity (Hofmann et al. 2009; O'Quin et al. 2010). These correlations suggest that opsin gene expression varies adaptively in cichlids, especially since some expression-ecology correlations have evolved independently among cichlids in different lakes (O'Quin et al. 2010). A recent quantitative genetic analysis of opsin expression in two Lake Malawi cichlids found a quantitative trait locus (QTL) located near the opsin genes (Carleton et al. 2010). The proximity of this QTL to the opsins suggests that mutations within one or more *cis*-regulatory sequences may contribute to variation in cichlid opsin expression. But like many non-model systems, few genomic resources are currently available for cichlids, making it difficult to identify potential *cis*-regulatory alleles and test their association with opsin gene expression.

Here, we sequence and analyze bacterial artificial chromosome (BAC) clones containing the opsin genes of two African cichlid species, *Oreochromis niloticus*

(Katagiri et al. 2001) and *Metriaclima zebra* (Di Palma et al. 2007). These clones provide the genomic resources necessary to identify putative *cis*-regulatory sequences surrounding the opsins genes. *Oreochromis niloticus* (the Nile tilapia) is a riverine cichlid that expresses the long wavelength-sensitive opsin palette as adults but also expresses the other palettes as fry and juveniles (Carleton et al. 2008). *O. niloticus* is an outgroup to the diverse haplochromine cichlids endemic to Lakes Tanganyika, Malawi, and Victoria. *Metriaclima zebra* (the 'classic' Zebra cichlid) is one such haplochromine cichlid found in Lake Malawi. *M. zebra* expresses the short wavelength-sensitive opsin palette as an adult and during all developmental stages (Carleton et al. 2008). Both species last shared a common ancestor ~ 18 MYA, whereas *M. zebra* diverged from other phenotypically diverse Lake Malawi cichlids less than 2 MYA (Genner et al. 2007). After sequencing the opsin-containing BAC clones from these species, we then performed phylogenetic footprinting to identify putative *cis*-regulatory sequences within ~ 30 kb windows of non-coding sequences surrounding the opsins. We searched the resulting conserved non-coding elements (CNEs) for binding sites of 12 transcription factors important for vertebrate opsin expression (Browman and Hawryshyn 1994a; Browman and Hawryshyn 1994b; Ng et al. 2001; Dann et al. 2004; Peng et al. 2005; Roberts et al. 2005; Applebury et al. 2007; Takechi et al. 2008) (Table 5-1). Among others, these transcription factors include cone-rod homeobox protein (CRX) (Takechi et al. 2008), thyroid hormone receptor (THR) (Ng et al. 2001), and retinoic acid receptor (Browman and Hawryshyn 1994a). We repeat the search of TFBS in the nearby proximal promoter of each opsin as well. Finally, we also perform an analogous search for microRNA

target sites within the 3'-UTR of each opsin, since miRNAs can also influence gene expression (Arora et al. 2007; Xu et al. 2007; Conte et al. 2010) and many UTR sequences exhibit significant divergence among cichlids (Loh et al. 2010). After performing these phylogenetic footprinting steps, we next perform phylogenetic shadowing by comparing the proportion of shared and divergent TFBS and microRNA target sites between *O. niloticus* and *M. zebra*. In each region we compare the number of divergent TFBS/ miRNA target sites with the number expected given the over-all sequence divergence of intronic sequences (a measure of neutral evolutionary divergence (Keightley and Gaffney 2003; Halligan et al. 2004). These comparisons are used to identify putative *cis*-regulatory sequences that have undergone significant evolutionary divergence between our two focal species. Finally, we repeat our phylogenetic shadowing analysis of proximal promoters and 3'- UTRs using a panel of 18 phenotypically diverse cichlids from Lake Malawi. This final analysis allows us to determine whether the trends we identify for *O. niloticus* and *M. zebra* are generally applicable to the more closely related cichlids of Lake Malawi.

We find that many non-coding regions are highly conserved between *O. niloticus* and *M. zebra*, as well as among the closely related cichlids of Lake Malawi. However, we find at least two CNEs, five proximal promoters, and two 3'-UTRs that exhibit significant divergence in the number and type of TFBS and microRNA targets identified in *O. niloticus* and *M. zebra*. We also identify at least three alleles that are weakly associated with *SWS2A*, *RH2B*, and *LWS* expression – three opsins that are strongly differentially expression among cichlid species. These results suggests that

cis-regulatory sequences may contribute to opsin expression differences among African cichlids, and provide numerous candidates for future functional studies.

Materials and Methods

Sequencing and Assembly of Opsin-Containing BAC clones

We isolated clones containing the opsin genes from BAC libraries of two African cichlids, *Oreochromis niloticus* (Katagiri et al. 2001) and *Metriaclima zebra* (Di Palma et al. 2007). For *O. niloticus,* we used PCR to screen pooled clones from the T3 and T4 libraries (Katagiri et al. 2001). Primers used for these screens were: *SWS1* (*F*: TACCTGCAGGCTGCCTTTAT; *R*: CTCGCATGGAGGCTAAGAAC), *RH2A* (*F*: GCAGACCCGATCTTCTTCAA; *R*: AGCAGACGTGATTGTGATGG), *LWS* (*F*: TCCTGTGCTACCTTGCTGTG; *R*: ACAACGACCATCCTGGAGAC). We first chose 10 super-pools, each covering 10% of the entire 35,000 pooled clones, and screened them for opsin-positive plates. We then screened row and column pools from the plates with positive results to identify the exact clones containing the opsins. Fingerprinted contigs (FPCs) corresponding to the positive clones were identified and all clones in the contig were PCR tested for the opsins (Figure S5-1 [Appendix 4]). Contig geometries were confirmed by end sequencing the BACs, designing primers, and PCR testing. Based on the resulting alignments, one clone for each opsin array was selected for sequencing.

DNA from the selected clones was prepared using the Qiagen® MaxiPrep Plasmid Purification kit following the manufacturer's protocols. The *O. niloticus* clones were sent to the Joint Genome Institute (JGI) for sequencing. Shotgun

libraries were prepared and 4 x 384-well plates were sequenced using ABI technology in both forward and reverse directions. The resulting reads were base-called and assembled with phred (Ewing et al. 1998) and phrap (Green 1994). Additional reads for the *SWS1*-containing clone were generated using 454 Life Sciences technology (Margulies et al. 2005). We performed two different sequencing runs for this clone, assembled them into contigs, and combined them with the JGI ABI reads in Sequencher v4.9 (Gene Codes Corporation, Inc.). This resulted in several large but non-overlapping contigs. To finish joining these contigs we used BLAST (Altschul et al. 1990) and Pipmaker (Schwartz et al. 2000) to identify and align the largest contigs to orthologous genomic regions from the genomes of other teleost fish (for an example see Figure S5-2 [Appendix 4]). Based on these alignments we designed PCR primers to sequence across the gaps to join the contigs.

For *M. zebra* we screened high-density BAC array filters using filter hybridization (Di Palma et al. 2007). This utilized PCR probes generated from *M. zebra* retinal cDNAs that were labeled using the ECL Nucleic Acid Labelling and Detection Kit (Amersham Biosciences). We obtained three clones from these arrays and confirmed that they contained the opsins via PCR as detailed above. DNA for these clones was prepared using the Qiagen® MaxiPrep kit following the manufactures protocols. BAC clones were sized by pulsed field gel electrophoresis following digestion with NotI. We then sent the purified, sized samples to 454 Life Sciences (Branford, CT) for sequencing. We performed two sequencing runs on the *SWS1* and *LWS*-containing clones, but only one for the clone containing *RH2A*. Due to the size of the 454 reads the resulting sequences formed more, but smaller contigs

relative to *O. niloticus*. To finish joining these contigs we aligned the largest (> 5 kb) contigs to the finished *O. niloticus* BAC sequences in Sequencher v4.9 and once again designed PCR primers to sequence across the gaps. We annotated the BAC sequences for both *O. niloticus* and *M. zebra* using BLAST (Altschul et al. 1990).

Finally, we performed a global alignment of each BAC from *O. niloticus* and *M. zebra* in the program wgVISTA (Couronne et al. 2003) (see Figure S5-3 for a Pip plot of each opsin-containing BAC from these two species). We measured sequence similarity and divergence across each BAC using the phylip program dnadist, implemented in the Mobyle online bioinformatics server (Néron et al. 2009). When measuring pairwise sequence divergence (D_{xy}), we used the Jukes-Cantor nucleotide model to correct for multiple hits. We repeated these measurements for each of the CNEs, promoter regions, and 3'-UTRs. We compared D_{xy} among each of these regions and the entire BAC sequences using t-tests implemented in the statistical software package R v2.10.0 (R Development Core Team 2009). Prior to performing all tests, we transformed the D_{xy} scores by $\log_{10}$ in order to meet the assumption of normality of errors.

Phylogenetic Analyses

We generated phylogenies of the teleost *RH2* and *SWS2* opsins in order to identify orthologous opsins among the focal fish genomes examined. We accessed all relevant opsin sequences from the genome assemblies of four fish genomes (see Methods for conserved non-coding elements below) via BLAT. We aligned both opsin data sets using the E-INS-i strategy of the multiple alignment program MAFFT v6.0 (Katoh and Toh 2008) and then chose an appropriate model of nucleotide

substitution via the program jModelTest v0.1.1 (Posada 2008). This model was TIM3ef+G for both the *RH2* and *SWS2* alignments. We then used this model and the corresponding parameters estimated by jModelTest to generate Neighbor-Joining trees for the opsins with Maximum Likelihood-corrected distances. For the *RH2*/*SWS2* datasets, these parameters included the nucleotide substitution rate matrix (A-C: 0.601/0.617; A-G: 1.470/1.734; A-T: 1.00/1.00; C-G: 0.601/0.617; C-T: 2.729/2.877; G-T: 0.599/0.155) and the shape of the gamma distribution (0.507/0.577). We measured the nodal support of these trees with 1000 bootstrap replicates. We rooted both trees using the *LWS-1* opsin of zebrafish.

Identification and Analysis of Conserved Non-coding Elements (CNEs)

We used phylogenetic footprinting (Gumucio et al. 1996) to identify putative *cis*-regulatory elements by searching for conserved non-coding elements (CNEs) surrounding the opsin gene arrays. To do this, we identified 100 – 300 kb regions of orthology between the *O. niloticus* BAC sequences and the genome assemblies of four teleosts fishes using BLAT and the UCSC genome browser. These additional genomes were stickleback (*Gasterosteus aculeatus*, Broad Institute v1.0, February 2006), medaka (*Oryzias latipes*, National Institute of Genetics and the University of Tokyo v1.0, October 2005), pufferfish (*Tetraodon nigroviridis*, Geoscope and Broad Institute v7, February 2004), and zebrafish (*Danio rerio*, Trust Sanger Institute zv8, December 2008). We then determined the gene locations of known opsin genes and examined synteny across these regions via DOT plots generated in the program PipMaker (Schwartz et al. 2000) (for an example, see Figure S5-2 [Appendix 4]). Regions of high synteny surrounding the opsins were then identified using

MultiPipMaker (Schwartz et al. 2003) (Figure 5-1). We defined a CNE as any region ≥ 50 bp long that was conserved between *O. niloticus* and at least one other teleost species (*O. latipes*, *G. aculeatus*, and *T. nigroviridis*). In each case, we attempted to analyze as many CNE as possible, but acknowledge that some small regions may have been missed.

Profiling Transcription Factor Binding Sites (TFBS)

We identified binding sites within each CNE as well as the proximal promoters located approximately 1 kb upstream of the opsin's translation start sites (TSS) using the Transcription Element Search System, TESS v6.0 (Schug 2008). We limited our search to high quality matches by accepting only those hits that met three criteria: (*i*) a log-likelihood (L_a) score ≥ 9.0, (*ii*) a ratio of the actual log-likelihood score to the maximum possible log-likelihood (L_q) score ≥ 80%, and (*iii*) a probability value for the log-likelihood score (L_{pv}) < 0.05. Although TESS can potentially identify binding sites for many different transcription factors, we were primarily interested in those factors that have been shown to influence opsin expression in fish and other vertebrates (Table 5-1). Following the automated search in TESS, we manually searched the lists for duplicate sites at each position, and removed them prior to further analysis.

Additionally, we also analyzed the number of shared and unique transcription factor binding sites found in each CNE and opsin proximal promoter from *O. niloticus* and *M. zebra*. We counted the total number of orthologous binding sites in both species, as well as those that were found in only one species or the other. These observed numbers of shared and divergent sites were compared to the expected

numbers suggested by the global sequence similarity of the *O. niloticus* and *M. zebra* BACs (92% versus 8%, see Results and Disscussion). We tested the independence between these observed and expected proportions using Binomial exact tests (Sokal and Rohlf 1995) implemented in the R statistical software package. To control the Type I error rate for each region examined, we calculated Bonferroni-corrected p-values for all test in R.

Finally, we also compared the average number of binding sites for each transcription factor between the proximal promoters of the *O. niloticus* opsins and seven randomly chosen, non-opsin genes from a draft assembly of the *O. niloticus* genome (available at www.BouillaBase.org; accessed October 2010). These genes were *ACTG1*, *AMPD3*, *DHCR7*, ENSGAC000000020282, *IGFALS*, *KCNJ9*, and *REEP1*. Proximal promoters from these randomly chosen sequences were identified based on comparison of the *O. niloticus* genes with orthologus regions from the stickleback genome. Comparison of the average number of binding sites across all opsins and transcription factors was performed using a Wilcoxon paired signed-rank test computed in R.

Comparison of Opsin Expression and TFBS profiles in Oreochromis niloticus

We evaluated the correlation between the transcription factor binding sites in the proximal promoter of each opsin and the expression of each opsin among developing *O. niloticus* fry using Mantel's test of two distance matrices. We generated Euclidean distance matrices of the total number of binding sites for 12 transcription factors within the proximal promoter region of each opsin as well as the percent of total opsin expression from developing *O. niloticus* fry, reported in

Carleton et al. (Carleton et al. 2008). We calculated Mantel's test using the 'mantel.randtest' function from the R package ade4 (Chessel et al. 2004). Approximate p-values were calculated following 500 randomizations of each matrix. We also expanded this analysis to the entire proximal promoter region after calculating a sequence similarity matrix for the entire proximal promoter using the phylip program dnadist.

Profiling of microRNA target sites

We searched the 3'-UTRs of each opsin for binding sites matching the target seed of known microRNAs (miRNA) via the SeedMatch algorithm previously developed to identify microRNA targets in cichlid UTRs (Loh et al. 2010). This algorithm is similar to the TargetScanS algorithm used in other studies (Lewis et al. 2005). Briefly, non-redundant fish miRNA targets were obtained from miRBase (www.mirbase.org (Griffiths-Jones et al. 2007); accessed June 2010) and supplemented with several miRNA target sequences identified in cichlids (Loh et al. 2010). We searched each opsin 3'-UTR—defined as the ~ 500 bp region between the transcription end site and the polyadenylation site (AATAAA)—for sequences matching the seeds of miRNAs from this non-redundant library. In order to account for the high rate of false-positive generated by simply searching for matching seed sites, we aligned the 3'-UTR of the cichlid opsins with those from *G. aculeatus*, *O. latipes*, *T. nigrovirdis*, the Japanese pufferfish (*Tetrapdon rubripes*), and *D. rerio* in order to identify sites that were conserved across multiple fish species. For this purpose we defined the first 1 kb of sequence downstream of these latter species' opsins as the 3'-UTR and aligned these to the cichlid sequences with MLagan

(Brudno et al. 2003). To account for errors in the alignment of orthologous 3'-UTRs, we counted as conserved the same miRNA target seed found within 50 bp of each other across species. For cichlid opsins that lacked orthologs in the other species, we used the nearest paralog (see Figure S5-4 [Appendix 4]).

Phylogenetic Shadowing of Opsin Promoter and 3'-UTR Sequences in Lake Malawi Cichlids

If the conserved non-coding sequences we identify do in fact represent functional *cis*-regulatory regions, then mutations in these sequences should be associated with altered opsin expression. We tested this hypothesis for the proximal promoters of five opsins, CNE 10, and two opsin 3'-UTRs. Using 18 phenotypically divergent Lake Malawi cichlid species, we sequenced approximately 1 kb of DNA upstream of the translation start site for 5 opsins, 0.9 kb around CNE 10, and approximately 0.5 kb downsream of *SWS2B* and *LWS*. We generated primers for these regions based on the *O. niloticus* and *M. zebra* BAC assemblies. The taxa sampled are listed in Table S5-1 [Appendix 4] along with their GenBank accession numbers; the primers used to generate these sequences are listed in Table S5-2 [Appendix 4]. We measured opsin expression for each individual following the protocols reported previously in Spady et al. (Spady et al. 2006) and Hofmann et al. (Hofmann et al. 2009).

Following sequencing, we estimated polymorphism statistics for the resulting sequences, and also performed a sliding-window analysis of nucleotide diversity (π), in the program DnaSP v5 (Librado and Rozas 2009). For the sliding-window analysis, we ignored all gaps and specified a window length of 50 and a step size of

10. Finally, we calculated the statistical association between polymorphisms found in CRX binding sites and peaks of nucleotide diversity among the sampled taxa using linear regression in the program gPLINK v1.07 (Purcell et al. 2007). For each test, we estimated the association of each locus with the expression of its downstream opsin, using membership in one of two major phylogenetic clades (mbuna and utaka; see Table S5-3 [Appendix 4]) as a covariable.

Results and Discussion

BAC Sequencing and Analysis

BAC Identification, Sequencing, and Assembly

Within the cichlid genome, the opsins are found in three separate tandem arrays. *SWS1* is found alone on cichlid linkage group (LG) 17; *SWS2A*, *SWS2B*, and *LWS* are found together in a tandem array on LG 5 (Lee et al. 2005); and *RH2B*, *RH2Aα*, and *RH2Aβ* are found in a second tandem array on LG 5 approximately 30 cM from the *SWS2-LWS* array (KL Carleton, unpublished data). We identified opsin-containing clones for *O. niloticus* by PCR screening its BAC library using primers for one opsin in each array. These clones were then shotgun sequenced using ABI Sanger sequencing technology (Table 5-2) and assembled in Sequencher v4.9 (Gene Codes Corporation, Inc.). The average read length for these sequences was ~700 bp. The final assembly of the *LWS*-containing clone generated a single contig of 171.8 kb, and assembly of the *RH2*-containing clone produced a final contig of 185.2 kb. Assembly of the *SWS1*-containing clone based on ABI-generated reads was poor, and we were able to assemble only 50% of the available reads. Additional 454 reads of

this clone were able to join all reads into a contig of 172.5 kb. The final assemblies of each clone joined > 84% of reads into a single contig that was within 10 – 40 kb of the estimated clone size (Table 5-2). All assemblies successfully covered the opsin-containing regions in *O. niloticus*.

For *M. zebra* we used filter hybridization to screen its BAC library and identify three clones containing the opsin arrays (Di Palma et al. 2007). These clones were then sequenced using 454 technology (Margulies et al. 2005). The average read length produced by this method was ~ 110 bp. Assembly of the *LWS*-containing clone produced a contig approximately 107.7 kb in length and covered the entire length of the *LWS*, *SWS2A*, and *SWS2B* opsin-containing region. Assembly of the *RH2*-containing clone produced two contigs of 29.39 and 48.02 kb. The alignment of these two contigs with *O. niloticus* suggests that their ends are approximately 500 bp apart, but several attempts to join these contigs using PCR failed. The gap between by these contigs is located between the *RH2Aα* and *RH2Aβ* opsins. These two genes are oriented away from each other and therefore share a common upstream promoter region (Figure 5-1). This gap may be the result of an inversion in *M. zebra* relative to *O. niloticus,* although the rest of the *M. zebra* sequences exhibited strong synteny with the *O. niloticus* sequences (Figure S5-3 [Appendix 4]). In all, we successfully covered 7.03 and 2.7 kb upstream of each *RH2A* gene, but we cannot be certain the entire promoter region of these two opsins was sequenced or assembled. Assembly of the *SWS1*-containing clone produced a contig of 77.65 kb. Except for the *RH2Aα* and *RH2Aβ* opsins, each assembly produced contigs very near in size to the estimated clone size (Table 5-2) and successfully covered the three opsin-containing regions in

M. zebra.

We aligned each BAC assembly from *O. niloticus* and *M. zebra* and found them to be highly similar. The average pairwise Jukes-Cantor-corrected sequence divergence (D_{xy}) across each BAC assembly was 8.4% (± 3.1% s.e.). This rate of sequence divergence is consistent with comparisons of other genes between these species, and it is one of the first large-scale estimates of sequence divergence between *O. niloticus* and *M. zebra*. We then further subdivided each BAC assembly into opsin protein-coding (CDS) and intronic (INT) sequences. For our two focal species, the mean D_{xy} across all opsin CDS was 3.8% (± 0.3%), while the divergence across all INT was 9.5% (± 1.9%). (We excluded both the first intron as well as the first and last six bases of each intron since these regions may contain regulatory sequences and splice sites that are more highly conserved than other intronic regions (Keightley and Gaffney 2003)). Comparison of the average D_{xy} across all regions shows that the mean divergence of the functionally important opsin CDS is significantly lower than D_{xy} across either the BACs or INT sequences (t-tests: CDS vs. BAC, $t_{8, 0.05} = 2.60$, $p = 0.032$; CDS vs. INT, $t_{27, 0.05} = 2.17$, $p = 0.039$), but that D_{xy} between BAC and INT sequences do not differ ($t_{23, 0.05} = 0.08$, $p = 0.935$). These results suggest that pairwise sequence divergence across all BAC sequences should provide a good estimate of neutral evolutionary divergence. These estimates also provide an important null hypothesis for our subsequent analyses using phylogenetic shadowing: in general, we expect *O. niloticus* and *M. zebra* to share (e.g, exhibit orthology in) ~ 92% of their TFBS and miRNA targets, and be divergent in ~ 8%. Divergence in greater than 8%

of the TFBS and miRNA target sites identified may indicate significant *cis*-regulatory evolution in the regions examined.

BAC Annotation and the Opsin Repertoire of Teleost Fishes

In order to perform phylogenetic footprinting across the opsin arrays of cichlids, we first investigated the synteny of each opsin array relative to several teleost fish species using PipMaker (Schwartz et al. 2000) and MultiPipMaker (Schwartz et al. 2000). We found considerable synteny in the opsin-containing regions among *O. niloticus* (tilapia), *Gasterosteus aculeatus* (stickleback), *Oryzias latipes* (medaka), *Tetraodon nigroviridis* (tetraodon), and *Danio rerio* (zebrafish) (Figure 5-1; Figure S5-2 [Appendix 4]). The clearest example of this synteny was the *SWS2-LWS* opsin array. This array is flanked by the genes *HCFC1* and *GNL3L* and is essentially co-linear in all five fish genomes (Figure 5-1; Figure S5-2 [Appendix 4]). We found evidence for a localized duplication of the *SWS2* opsins in *O. latipes* and *O. niloticus*, since both these species have two adjacent *SWS2* opsin genes. Closely related Poeciliid fishes also possess adjacent *SWS2* paralogs (Watson et al. 2010), suggesting that this duplication event probably occurred at least 153 - 113 MYA at the base of the Acanthopterygii (Carleton and Kocher 2001; Steinke et al. 2006).

In contrast to the *SWS2-LWS* array, we observed considerable variation in opsin gene content for the *RH2* opsins. *O. niloticus* and *M. zebra* possess three *RH2* genes while *D. rerio* has four (Chinen et al. 2003; Hofmann and Carleton 2009), *G. aculeatus* has two, and *T. nigrovirdis* has one functional *RH2* opsin and one *RH2* pseudogene (Neafsey and Hartl 2005). We therefore used phylogenetic analyses to investigate the orthology of the *RH2* and *SWS2* genes among these fishes and found

that most *RH2* duplications are species-specific (Yokoyama and Tada 2010) (Figure S5-4 [Appendix 4]). Thus, synteny in the region containing the *RH2* opsin array was lower than in the *SWS2-LWS* array, but was still largely co-linear between *O. niloticus*, *G. aculeatus*, and *T. nigroviridis* (Figure S5-2 [Appendix 4]). The genes *SLCA613-like* and *SYNPR* flank the *RH2* opsins in these fishes (Figure S5-5).

Synteny in the region surrounding the *SWS1* opsin was also difficult to assess due to species-specific deletions and poor sequence assembly. The *T. nigrovirdis* genome assembly lacks the *SWS1* opsin altogether, and this region is poorly-assembled in the *G. aculeatus* and *O. latipes* genomes. *SWS1* is found within an unordered chromosome or ultracontig in both species, but appears to be flanked by the genes *TNPO3* and *SOCS2* (Figure S5-5). For *G. aculeatus*, we found a small 92 kb region containing the *SWS1* opsin that was collinear with the *O. niloticus* BAC sequence, but which contained one large inversion. For *O. latipes*, we found an even smaller 60 kb region that was syntenic for only 11 kb surrounding the *SWS1* opsin. Synteny with *D. rerio* was also generally low (Figure S5-2 [Appendix 4]). Therefore, despite the lack of duplicates of the *SWS1* opsin, this region is generally poorly assembled in the existing annotations of several teleost genomes, complicating direct comparisons of synteny in this region.

A diagram of the opsin arrays for all five species examined is shown in Figure S5-5 [Appendix 4]. Despite several instances of gene duplication and local sequence divergence, these differences at most represent small perturbations within larger regions of high conservation and synteny.

Analysis of Conserved Non-coding Elements (CNEs)

Phylogenetic Footprinting of CNEs

We used MultiPipMaker (Schwartz et al. 2003) to highlight non-coding elements surrounding each opsin gene array from *O. niloticus* to *D. rerio*, representing nearly 300 MY of fish evolution (Steinke et al. 2006). The resulting plots illustrate at least 23 conserved non-coding elements (CNEs) surrounding the opsins arrays of *O. niloticus* and the other species examined (red bars in Figure 5-1). We also found six regions of putatively high conservation that are largely composed of repetitive sequence, including one potential LTR transposon (green bars in Figure 5-1), which we did not analyze further. The conservation of these CNEs over several million years of fish evolution suggests that they contain functionally important regulatory modules necessary for gene expression.

At least one CNE we identified through phylogenetic footprinting is orthologous to other vertebrate *cis*-regulatory sequences. CNE 10 (highlighted in Figure 5-1 and located between the *SWS2B* and *LWS* opsins) consists of two non-contiguous regions of high conservation in pufferfish, stickleback, medaka, swordtails, and cichlids (Watson et al. 2010) (Figure 5-1). Each region is ~ 100 bp long, and they are separated by ~ 440 bp in *O. niloticus*. The first region, CNE 10a, was also identified following a comparative analysis of opsin-containing BACs from swordtails (*Xiphophorus helleri*) (Watson et al. 2010). Through BLAST and mirbase (Griffiths-Jones et al. 2007), we found that CNE 10a is most similar to zebrafish microRNA dre-miR-726 (score 173.3, e-value = 0.006), and the same genomic region from zebrafish is identical to this microRNA (Figure 5-2). Dre-miR-726 is expressed

in the retina of larval and adult zebrafish (Kloosterman et al. 2006). Since many microRNAs are transcribed along with the genes they regulate, the proximity of this microRNA to the *SWS2* and *LWS* opsins suggests that mir-726 could play a role in opsin regulation. The ~ 90 bp CNE encoding mir-726 is conserved in numerous other taxa as well, including additional fishes, frogs, and lizards (GenomeWiki 2009; Watson et al. 2010).

The second highly conserved region, CNS 10b, is positionally and structurally orthologous to the mammalian LWS locus control region (LWS-LCR; Figure 5-2) (Wakefield et al. 2008; GenomeWiki 2009; Watson et al. 2010). This enhancer is located ~ 3.8 kb upstream of the *LWS* opsin in *O. niloticus* and other vertebrates, including humans. The LWS-LCR is hypothesized to enhance *LWS* expression in eutherian mammals by looping and binding to the *LWS* proximal promoter (Wang et al. 1992; Smallwood et al. 2002; Wakefield et al. 2008). Wang et al. (Wang et al. 1992) demonstrated that the human ortholog of this sequence can function as an enhancer of both *LWS* and *MWS* (*SWS2*) opsin expression in mice, and Wakefield et al. (Wakefield et al. 2008) hypothesized that it could function as a bidirectional enhancer of both *LWS* and *SWS2* expression in numerous other species as well. Earlier surveys of *cis*-regulatory sequence surrounding the opsins of zebrafish (Takechi et al. 2008) did not find an LCR homolog between the *SWS2* and *LWS* genes; however, a recent analysis of the *LWS* opsin array in green swordtails (Watson et al. 2010) suggest that a homologous LWS-LCR sequence is present in *D. rerio* and other teleosts, including cichlids. Although we cannot conclude that the LWS-LCR homolog we identify actually regulates opsin expression in fishes, the identification

of this functionally important region from mammals is encouraging. Our results demonstrate the effectiveness of the phylogenetic footprinting method we use for identifying functional *cis*-regulatory sequences necessary for vertebrate opsin expression. It is possible that the remainder of these CNEs also encode *cis*-regulatory sequences necessary for the correct spatial and developmental expression of the opsins in cichlids.

We note that our present study focuses on small regions of high conservation within an ~ 30 kb window of non-coding sequence surrounding the opsin arrays, but that *cis*-regulatory sequences may often reside tens or hundreds of kilobases from the genes they regulate. However, two recent analyses of general transcription factor binding sites found that functional binding sites generally cluster in regions 1 kb around the proximal promoter of each gene (Mann and Carroll 2002; Birney et al. 2007; Rozowsky et al. 2009). This observation suggests that a focused study of conserved elements within or near the opsins is a reasonable strategy for this initial study. A FASTA file of all CNE sequences from *O. niloticus* and *M. zebra* is provided in Supplementary File 4-1 [Appendix 4].

TFBS Profiling and Phylogenetic Shadowing of CNEs

We compared the 23 CNEs identified between *O. niloticus* and *M. zebra* and found many to be highly conserved; however, we could find no identifiable orthologs between *O. niloticus* and *M. zebra* for CNEs 9 or 22. CNE 22 is found between the two *RH2A* opsins and is likely included in the missing/unassembled region of the *M. zebra RH2*-BAC (see above). For the remaining orthologous CNEs, the average pairwise sequence divergence between *O. niloticus* and *M. zebra* was 4.7% (± 0.6%),

which is significantly less than the mean D_{xy} of introns (9.5%, t-test: $t_{41, 0.05} = 2.74$, $p = 0.009$). This result suggests that the conserved non-coding regions identified among *O. niloticus* and other fishes have remained conserved among African cichlids as well.

We used the Transcription Element Search System (Schug 2008) to computationally search all orthologous CNEs for binding sites corresponding to twelve transcription factors that have been associated with opsin expression in fishes and other vertebrates. These twelve transcription factors have been shown to influence opsin expression in many fish and vertebrate system (Schule et al. 1991; Salbert et al. 1993; Browman and Hawryshyn 1994a; Browman and Hawryshyn 1994b; Dann et al. 2004; Peng et al. 2005; Applebury et al. 2007; Takechi et al. 2008), and a complete list of these transcription factors and their associated opsins is presented in Table 5-1. We found computationally-predicted binding sites for these functionally important transcription factors in nearly all (22 of 23) CNEs surveyed (Table 5-3; see also Table S5-4 [Appendix 4] for detailed counts of all TFBSs). Only CNE 13 lacked binding sites for the any of the twelve transcription factors in either species examined. Within the remaining sequences we found binding sites for all twelve transcription factors except PNR and RXRγ. After relaxing our matching criteria, we still failed to find binding sites for these two transcription factors (data not shown). In both *O. niloticus* and *M. zebra*, binding sites for AP-1 and CRX were extremely abundant, although binding sites for each of three retinoic acid receptors (RARs) and THRβ were also common (Table S5-4 [Appendix 4]). The CNEs with the highest density of transcription factor binding sites (defined as the total number of

binding sites divided by total length surveyed – generally 9 TFBS or more; see Table S5-4 [Appendix 4]) include CNEs 2, 5, 6, 16, 22, and 23 for *O. niloticus*, and 2, 3, 5, 14, 16, and 23 for *M. zebra*. Due to their potential enrichment for functional TFBS relative to other CNEs, we believe these seven CNEs represent the most likely candidates for functional *cis*-regulators of opsin expression in fishes.

Consistent with the high similarity of their sequences, the results of our TFBS search varied very little between *O. niloticus* and *M. zebra*. We used exact binomial tests to compare the proportion of shared and divergent TFBS observed between *O. niloticus* and *M. zebra* to the theoretical ratio of 92:8% (see above). Treating each TFBS independently, we counted each non-orthologous TFBS as a success, each orthologus TFBS as a failure, then tested the hypothesis that the true probability of success (proportion of divergent TFBS, P_{div}) was 8%. We found that only 2 of 20 CNEs differed significantly from this null expectation (Table 5-3). These are CNEs 3 and 6 (exact binomial tests: CNE 3, estimated proportion of divergence, $P_{div} = 42.9\%$, $p = 3.24e^{-4}$; CNE 6, $P_{div} = 77.8\%$, $p = 1.31e^{-5}$). These results did not change when we used the mean similarity of introns from each CNE's nearest down-stream opsin as a null hypothesis, except that CNE 7 also showed significant TFBS divergence ($P_{div} = 100.0\%$, $p = 0.018$). CNE 3 is located downstream of the *SWS1* opsin and CNEs 6 and 7 are located upstream of the *SWS2A* opsin. For CNE 3, *M. zebra* has more than double the number of TFBS than *O. niloticus*, including additional RARβ and RARγ binding sites. For CNE 6, *O. niloticus* has 8 TFBS while *M. zebra* has only one; and for CNE 7, *M. zebra* has two while *O. niloticus* has none. These results are consistent with what one might expect based on the expression of these opsins, since *SWS1* is

highly expressed in *M. zebra* but not *O. niloticus*, and the opposite is true for *SWS2A*. Thus, we show that *O. niloticus* and *M. zebra* have diverged significantly in the identity of their TFBS profiles for three putative *cis*- regulatory elements, and differ in the presence/absence of two more. These results offer the intriguing possibility that at least some of the differences in opsin expression observed between *O. niloticus* and *M. zebra* could be due to divergence in the TFBS profiles of CNEs surrounding their opsins.

We acknowledge that our use of the overall proportion of divergent TFBS (P_{div}) to detect CNEs that have undergone significant *cis*-regulatory divergence ignores many nuances in TFBS divergence, such as the overall number and kind of TFBS present in each CNE and species. But because of the small number of TFBS found within each CNE (the average number of TFBS found in each CNE is < 6), it is difficult to perform robust tests of divergence in individual TFBS. Therefore, we have summed all TFBS into orthologous (shared) and non-orthologous (divergent) groups in order to measure TFBS divergence between *O. niloticus* and *M. zebra*. However, even within these broad categories, we have only enough power that CNEs with $P_{div} > 40\%$ stand out as statistical outliers. In the future we aim to perform more nuanced, sequenced-based tests of *cis*-regulatory divergence in cichlids, but only after functional CNEs controlling opsin expression are identified. We present these tests for *cis*-regulatory divergence as a first step in that process.

Analysis of Proximal Promoter Regions

Phylogenetic Footprinting of Opsin Proximal Promoters

The MultiPip plots shown in Figure 5-1 reveal 23 CNEs upstream of the opsins, but also show several regions of high conservation within the 5' proximal promoter of multiple opsins as well. In particular, *SWS2A*, *SWS2B*, and *LWS* all exhibit regions of strong conservation in the first 1 kb of sequence upstream of their translation start site (TSS). For the *LWS* opsin, this region of conservation spans nearly the entire proximal promoter (~ 0.7 kb) and multiple fish species, including *G. aculeatus*, *O. latipes*, and *T. nigroviridis* (Fig. 4-1B). *RH2A*α and *RH2A*β also exhibit some small regions of high conservation just upstream of their TSSs, which probably reflect the 5'-UTR region. *RH2B* also contains some conserved regions, though these are largely composed of repetitive sequences (Fig. 4-1C). It is intriguing that many of the opsins exhibit strong conservation of elements within 1 kb of their TSS, which we use to define the proximal promoter, because the true promoter regions for these genes are unknown in cichlids. But important *cis*-regulatory sequences have been identified in close proximity to the opsin genes in other fish species. In particular, several CRX transcription factor binding sites found within 500 bp of the *SWS2* opsin that regulate the expression of this gene in *D. rerio* (Takechi, Seno, Kawamura 2008). Therefore, the conservation we observe upstream of the *SWS2A*, *SWS2B*, and *LWS* opsins may indicate the presence of additional *cis*-regulatory sequences with the proximal promoters of these genes as well.

TFBS Profiling and Phylogenetic Shadowing of Opsin Proximal Promoters

The distribution and number of transcription factor binding sites found within the proximal promoter sequences of each opsin was similar for those found in the CNEs. Within each opsin's proximal promoter region, AP-1 and CRX binding sites

were nearly ubiquitous (Figure 5-3). Binding sites for NFκB, RARα, RARβ, RXRβ and THRβ were also common, and we once again found no binding sites for PNR and RXRγ. The absence of binding sites for PNR and RXRγ in both the CNEs and promoters may rule-out these factors as candidate *trans*-regulators of cichlid opsin expression; however the lack of these factors could also be due to biases in the way TESS identifies binding sites. Interestingly, we found several CRX binding sites directly upstream of the *SWS2A* and *SWS2B* opsins (Figure 5-3). These binding sites could potentially function as regulators of blue opsin expression in cichlids as they do in zebrafish (Takechi et al. 2008).

Pairwise sequence divergence in the proximal promoter regions was greater than for the other regions examined. The average D_{xy} of the proximal promoters was 10.2% (± 3.2%), which differed significantly from the mean of CNEs (4.7%, t-test: $t_{26, 0.05} = 2.20$, $p = 0.037$), but not the introns (9.5%, t-test: $t_{27, 0.05} = 0.14$, $p = 0.89$). This result suggests that the opsin promoter regions of cichlids may exhibit greater divergence in putative *cis*-regulatory sequences than the CNEs. Indeed, we found that *O. niloticus* and *M. zebra* exhibited significant divergence in their TFBS profiles for five of the seven opsin proximal promoters (Figure 5-3). Following correction for multiple hypothesis testing, *O. niloticus* and *M. zebra* exhibited significant differences in their TFBS profiles for the *SWS1*, *SWS2A*, *RH2B*, *RH2A*α and *RH2A*β opsins (Figure 5-3; see also Table 5-3). *O. niloticus* and *M. zebra* differ dramatically in their adult expression of all of these opsins (Carleton et al. 2008), suggesting that their divergent transcription factor profiles could explain these differences. A quick comparison of which TFBS differ between *O. niloticus* and *M. zebra* does not reveal

any extreme differences, other than a slight over-representation of CRX sites in *O. niloticus* (17 vs. 7), and of THRα in *M. zebra* (4 vs. 0) (Figure 5-3).

Using phylogenetic shadowing, we identified five cichlid opsins with promoter sequences that exhibit significant divergence in their binding site profiles for 12 transcription factors. We note, however, that by focusing on only these few TFBS, we potentially miss many interesting patterns of divergence in transcription factors that have not already been associated with vertebrate opsin expression. A comprehensive search of all TFBS identified by TESS could potentially pick up these missed patterns of divergence, but such a search would be extremely cumbersome and subject to many false positives (Wasserman and Sandelin 2004). Because of their small size, TFBS motifs are likely to appear throughout the genome frequently by chance, and it is difficult to determine which are likely to be functional based on sequence matches alone. Therefore, we opted to focus on genes that are obvious candidates for analysis. We performed an additional analysis to determine the relevance of these twelve candidates factors by comparing the number of TFBS of each factor found within the opsin proximal promoters and the proximal promoters of seven randomly chosen, non-opsin genes. We hypothesized that if these candidates are relevant to the control of opsin expression in cichlids (and thus to our analysis of *cis*-regulatory sequence divergence), then we would find a significantly greater number of TFBS for each factor upstream of the opsin genes compared to the non-opsin genes. Indeed, we found that the opsins contain a greater number of binding sites for eight out of ten factors compared to the non-opsin genes (Wilcoxon paired signed-rank test: $V = 50$, $p = 0.0124$; Figure 5-3H). The non-opsin promoters

contained higher mean numbers of TFBSs for AP-1 and THRβ only. This result suggests that the proximal promoters of the opsins are significantly enriched for the binding sites of transcription factors that influence vertebrate opsin expression. This enrichment also suggests that polymorphisms in these regions could conceivably lead to functional differences in transcription factor binding and opsin expression. A FASTA file of all opsin and non-opsin promoter sequences from *O. niloticus* and *M. zebra* is available in the supplementary File S4-1 [Appendix 4].

Finally, the opsins of cichlids are co-expressed together in three main groups, or palettes. These palettes generate visual pigment sets sensitive to short, middle, and long-wavelength spectra (Carleton et al. 2006; Hofmann et al. 2009; O'Quin et al. 2010). Therefore, as a final tangential analysis, we used the number and identity of predicted TFBS found upstream of each opsin to test the hypothesis that patterns of co-expression between functionally linked opsins are due to the presence of shared transcription factor binding sites in their promoters (Pennacchio and Rubin 2001). To test this hypothesis, we compared the opsin expression profiles of developing *O. niloticus* fry (Carleton et al. 2008) with the transcription factor binding site profiles from the proximal promoters of these same opsins. We chose developing *O. niloticus* fry because *O. niloticus* express all seven opsins over the course of development from 0 – 300 days post fertilization (Carleton et al. 2008). This developmental variation corresponds to the expression of each of the three opsin expression palettes observed among adult Lake Malawi cichlids (Carleton et al. 2008; O'Quin et al. 2011). We generated two matrices of Euclidean distances among the opsins based on their patterns of co-expression and the total number of TFBS identified for each candidate

transcription factor. We found no significant correlation between these expression and transcription factor binding site dissimilarity matrices (Mantel test: $r = -0.115$, $p = 0.623$). When we expanded the comparison to sequence similarity across the proximal promoter, the correlation did not improve but was actually reduced (Mantel test: $r = -0.036$, $p = 0.621$). Therefore, we conclude that the composite transcription factor binding site profiles and sequence similarity scores we record do not predict which opsins are co-expressed together in developing *O. niloticus* fry. It is possible that other transcription factors are responsible for the patterns of opsin co-expression observed in cichlid species.

The search parameters we have chosen aim to identify TFBS with high confidence while still accounting for the observation that many transcription factors exhibit degenerate binding of DNA motifs (Letovsky and Dynan 1989; Stormo 2000), and can bind these motifs in an orientation-independent manner (Baker et al. 1996; Latchman 2004). We are currently performing a quantitative genetic analysis of many markers located across the genome in order to identify other loci and transcription factors that may be associated with cichlid opsin expression. But even our preliminary analyses reveal that the divergent binding of transcription factors to several opsin promoters, particularly of CRX and THRα, could contribute to the differences in opsin expression observed among *O. niloticus* and *M. zebra*.

Analysis of Opsin 3' Untranslated Regions (3'-UTRs)

Phylogenetic Footprinting of Opsin 3'-UTRs

In addition to mutations within conserved non-coding elements and 5' promoter regions, polymorphisms within 3'-UTRs can also act as *cis*-regulatory alleles (Kloc et al. 2000; Chen and Rajewsky 2006). These polymorphisms affect gene expression by altering the binding of microRNAs (miRNAs) in a manner analogous to how mutations with TFBS can alter gene expression, except that miRNAs inhibit gene expression instead of promote it and this inhibition occurs post-transcriptionally.

Our phylogenetic footprinting analysis reveals that every opsin exhibits some conservation of the 50 – 100 bp region found directly downstream of the opsin coding sequences (Figure 5-1). Generally, this conservation is strongest between *O. niloticus*, *O. latipes*, and *G. aculeatus*, reflecting the close phylogenetic relationships of these species. For *RH2A*α, the 3' conserved region extends nearly 700 bp pass the end of the coding region. Initially, these results suggest that the opsin 3'-UTRs of cichlids will be highly conserved, reflecting the strong evolutionary constraint on UTR sequence and function seen in both flies and humans (Andolfatto 2005; Chen and Rajewsky 2006). However, many miRNAs are transcribed along with the genes they target, and our identification of miR-726 just upstream of the *LWS* opsin (see Figures 5-1 and 5-2) suggest that miRNAs could play an important role in regulating vertebrate opsin expression. Additionally, a recent survey of polymorphisms affecting microRNA target sites in cichlids suggests that these regions may in fact be under divergent selection in some species (Loh et al. 2010). Therefore, it is plausible that polymorphisms in microRNA target sequences could alter microRNA binding, and hence opsin expression, in African cichlids.

MicroRNA Target Site Profiling and Phylogenetic Shadowing of 3'-UTRs

We searched the 3'-UTRs of each opsin in *O. niloticus* and *M. zebra* for target sites corresponding to known fish microRNAs (Griffiths-Jones et al. 2007). In all, we identified 84 predicted target sites matching 30 known miRNAs from cichlids and *D. rerio* (Table S5-5 [Appendix 4]). Focusing on only those target sites that were conserved between African cichlids and other fish species, we found at least one conserved miRNA target site down-stream of each opsin (Table 5-4). Many of these conserved sites are expressed within the retina of vertebrates and play a role in retinal development (Wienholds et al. 2005; Ryan et al. 2006; Arora et al. 2007; Xu et al. 2007). For example, dre-miR-217, dre-miR-181a, and dre-miR-23b are integral to the development and maintenance of the vertebrate retina (Guerin et al. 2006; Li et al. 2007; Kato et al. 2009), while dre-miR-96 and dre- miR-182a are sensory organ-specific (Xu et al. 2007). Only one conserved site that was found in cichlids and other fishes differed between *O. niloticus* and *M. zebra*. A target site for dre-miR-722, found downstream of the *LWS* opsin in *O. niloticus* and the pufferfish (*Takifugu rubripes*), is missing in the orthologous 3'-UTR from *M. zebra* due to a single nucleotide polymorphism (SNP). However, the two conserved target sites for dre-miR-722 and dre-miR-728 are both found within the 3'-UTRs of several Lake Victorian cichlids (data not shown). Like *O. niloticus*, Lake Victoria's cichlids express the long wavelength opsin palette as adults (Terai et al. 2006), possibly indicating that these factors play a role in *LWS* expression. Additionally, the target sites we found within the *RH2A*α and *RH2A*β UTRs were identical, consistent with the short amount of time that has passed since these opsins were duplicated in cichlids

approximately 27 MYA (Parry et al. 2005; Yokoyama and Tada 2010). Therefore, if we interpret evolutionary conservation as an indication of functionality, we believe the conserved sites listed in Table 5-4 represent those miRNA target sites that are most likely to influence opsin expression in African cichlids.

Like the CNEs identified earlier, all 3'-UTR sequences generally exhibited high similarity between *O. niloticus* and *M zebra*. The average pairwise divergence (D_{xy}) for *O. niloticus* and *M. zebra* 3'UTRs was 5.2% (± 1% s.e.). This small level of divergence is very similar to the level observed for opsin coding sequences, though it did not differ from the average D_{xy} of introns (9.5%, t-test: $t_{27, 0.05} = 1.33$, $p = 0.196$). Consequently, the results of our miRNA target search were once again very similar for *O. niloticus* and *M. zebra*, especially for those sites conserved in other fishes as well (Table S5-5 [Appendix 4]). However, we also identified numerous sites that were not conserved between cichlids and other fishes, or between *O. niloticus* and *M. zebra*. In particular, the UTRs downstream of the *RH2B* and *SWS2B* opsins exhibited significant divergence in their number of orthologus and shared miRNA target sites (exact binomial tests: *RH2B*, estimated $P_{div} = 50.0\%$, $p = 0.015$; *SWS2B*, estimated $P_{div} = 66.67\%$, $p = 0.001$; see Table 5-3). These results did not change when we altered the null hypothesis to reflect the divergence of each opsin's intronic sequences (data not shown). For *RH2B*, we identified six distinct miRNA target sites, only two of which were shared between *O. niloticus* and *M. zebra* (dre-miR-135 and dre-miR-190); *M. zebra* then exhibited four additional target sites for miRs 101, 144, 196, and 2184 that *O. niloticus* lacked. For *SWS2B*, the two species exhibited orthologous target sites for miR-217, but *M. zebra* had additional targets for miRs 194 and 23,

while *O. niloticus* had additional sites for miRs 92 and 137. Thus, we not only identified conserved and perhaps core miRNA target sites within the 3'-UTRs of the opsins in cichlids, but we also identified divergent miRNA target profiles between *O. niloticus* and *M. zebra* for the *RH2B* and *SWS2B* opsins. *RH2B* is strongly differentially expressed in these two species, while *SWS2B* is only expressed in some adults of *O. niloticus* (Lisney et al. 2010). Finally, it is interesting to note that many of the conserved and non-conserved miRNA target sites we identify correspond to microRNAs associated with retinal development (for example, dre-miRs 23, 92, 722, and 194) (Calissano et al. 2007; Xu et al. 2007; Decembrini et al. 2009). Additionally, miR-129 is also associated with retinoblastoma in humans (Zhao et al. 2009). Given that *O. niloticus* and *M. zebra* differ dramatically in their developmental patterns of opsin gene expression, it is interesting to speculate that these miRNAs could contribute to the developmental differences in opsin expression observed between these and other African cichlid species (Carleton et al. 2008; O'Quin et al. 2011).

Whether conserved on not, we note that most miRNA target sites we identified correspond to miRNAs that are also expressed in the vertebrate retina (Table S5-5 [Appendix 4]). Of sites corresponding to 30 different microRNAs, 22 (73%) correspond to microRNAs expressed within the retinas of fish, mammals, or amphibians (Table S5-5 [Appendix 4]). Notably, however, we did not find any microRNA target sites that correspond to miR-726, the microRNA found upstream of the *LWS* opsin and encoded by CNE 10a (see Figure 5-2). Additionally, miRNAs may regulate gene expression by binding both to sequences within nascent mRNAs or

to their 3'-UTR. Here we have focused on target binding sites within the 3'-UTR of the cichlid opsins, but miRNA cleavage of messenger RNAs by binding to sites within core mRNA sequences has been demonstrated in both humans and plants (Hutvagner and Zamore 2002; Llave et al. 2002). It is still not clear which mechanism of miRNA regulation is more common, although a review by Bartel (Bartel 2004) suggested that translational repression by binding to UTR sequences is more predominant. Finally, we note also that the cellular machinery cannot distinguish between functional and non-functional miRNA target sites based on their evolutionary conservation in other species, as we do here (Bartel 2004). However, given that scans for miRNA target sites can have a high rate of false positives, evolutionary conservation is currently the best way to avoid high error rates and to infer functionality. The fact that we identified a high percentage of target sites that correspond to miRNAs found within the vertebrate eye suggests that many of these sites are not false-positives; therefore, it is plausible that they may actually function to regulate opsin expression in cichlids. In the future we will determine whether these and other miRNAs are actually expressed in the retinas of African cichlids. If so, then heterologous reporter assays could be used to verify what role interspecific differences in miRNA target sites may play in the evolution of cichlid opsin expression (Lewis et al. 2003; Stark et al. 2003). The sequences of all *O. niloticus* and *M. zebra* opsin 3'-UTRs are available in Supplementary File S4-1 [Appendix 4].

Phylogenetic Shadowing Among the Cichlids of Lake Malawi

Two goals of this study have been to (*i*) identify potential *cis*-regulatory sequences surrounding the opsin gene arrays of African cichlids, and (*ii*) identify

those putative regulatory sequences that may be undergoing divergent evolution among African cichlids with different patterns of opsin gene expression. For both of these goals we have relied on the BAC clones of *Oreochromis niloticus* and *Metriaclima zebra*—two species that have BAC clones available, but that also differ dramatically in their evolutionary age (~ 18 MY (Genner et al. 2007)) and adult and developmental patterns of opsin expression (Carleton et al. 2008). As a final goal, we wanted to determine whether the candidate *cis*-regulatory sites we identify also vary among a more closely related (~ 2 MY (Genner et al. 2007)) panel of cichlids from Lake Malawi. Although much more closely related to *M. zebra* than *O. niloticus*, adults of many Lake Malawi cichlid species exhibit the same opsin expression patterns as adult and juvenile *Oreochromis niloticus* (Carleton et al. 2006; Carleton 2009; Hofmann et al. 2009). Therefore, we re-sequenced some of the most promising candidate regions in a panel of 18 Lake Malawi cichlid species (Tables 5-5 and S5-3 [Appendix 4]). This panel included six species for each of the three adult opsin expression palettes observed among Lake Malawi's cichlids. The regions we re-sequenced included the proximal promoter regions upstream of five opsins (*SWS1*, *SWS2A*, *SWS2B*, *RH2B*, and *LWS*; highlighted in blue in Figure 5-1), CNE 10, and the 3'-UTR of the *SWS2B* and *LWS* opsins. After sequencing, we examined these regions for levels of polymorphism and performed a test of association for *cis*-regulatory alleles. Our panel of cichlids included 18 species that have been previously sampled from Lake Malawi; therefore, we knew *a priori* whether each species expressed the short, middle, or long wavelength-sensitive opsin palette. As a first step to this analysis, however, we confirmed these gene expression groups by measuring the

expression of each opsin in all individuals via RT-qPCR (Table S5-3 [Appendix 4]). Our opsin expression results were highly concordant with previous measurements (Hofmann et al. 2009).

We were able to sequence most of each proximal promoter in our panel of 18 Lake Malawi cichlids. We successfully sequenced the entire 1 kb region upstream of both the *SWS1* and *SWS2A* opsins. Additionally, we sequenced 956 bp upstream of *LWS*, 951 bp upstream of *RH2B*, but only 694 bp upstream of *SWS2B*. For the *SWS2A* proximal promoters we were only able to sequence 17 individuals. Finally, we sequenced ~ 450 bp downstream of the *SWS2B* and *LWS* opsins in 14 and 13 individuals, respectively. As expected given the young age of Lake Malawi cichlids, all putative promoter and 3'UTR sequences were highly conserved in the taxa sampled. In general, we identified fewer than 15 single nucleotide polymorphisms (SNPs) and insertion/deletions (indels) per region examined (Table 5-5). In each case, most SNPs were present as singletons or found in only one individual. Other diversity statistics—including the total number of segregating sites (S), total number of singletons (s), number of haplotypes (H), nucleotide diversity (π), sequence conservation (C), and Tajima's D (T_D)—also indicate overall low levels of polymorphism, despite our use of alternate species and genera as sampling units (see Table S5-3 [Appendix 4] for a list of all polymorphisms found among the 18 species sampled). Nevertheless, following a sliding window analysis of nucleotide diversity (π) and minor allele frequency (MAF), we were able to identify several peaks of relatively high π and MAF within each region examined (Figure 5-4). These peaks correspond to SNPs and indels segregating at high frequency within the species and

genera sampled, and therefore represent potential *cis*-regulatory alleles.

Several peaks of relatively high nucleotide diversity and MAF correspond to predicted TFBS. One peak located ~ 220 bp upstream of the *SWS2A* translation start site (TSS) contains one SNP and one 8 bp indel that both disrupt putative CRX binding sites. The 8 bp indel located at *SWS2A* -217 completely eliminates the CRX binding site in several taxa (Table S5-3 [Appendix 4]). We identified at least three other polymorphisms upstream of *SWS1* and *RH2B* that also disrupt CRX binding sites—all three present in only one individual—but no SNPs or indels interrupting the binding sites of the other candidate transcription factors (Table 5-5). Additionally, one peak of nucleotide diversity within CNE 10 corresponds to a SNP within the sequence of mir-726; however, this mutation does not occur within the mature miRNA. Finally, we found only three polymorphisms total within the 3'-UTRs of *SWS2B* and *LWS* (Figure 5-4; Table S5-3 [Appendix 4]). None of these three SNPs interrupted predicted miRNA target sites, in contrast to the polymorphism that segregates within the *LWS* 3-'UTR of *O. niloticus* and *M. zebra*. However, since mutations within transcription factor binding sites have been shown to alter gene expression (Kasowski et al. 2010), our results suggest that polymorphisms within the *SWS2A* promoter could contribute to the differential patterns of opsin gene expression observed among Lake Malawi cichlids.

To test this hypothesis, we performed allelic association tests between these and other SNPs underlying peaks of nucleotide diversity and high MAF (see Figure 5-4) with the expression of their nearest downstream opsin (Table 5-6). Three polymorphisms (*SWS2A*-217, *RH2B*-161, CNE10 570) exhibited significant or

marginally non-significant associations with the expression of their downstream opsins; however only *RH2B*-161 is significant following Bonferroni-correction for multiple hypothesis tests (Bonferroni cut-off: $\alpha = 0.05/11 = 0.0045$; *RH2B*-161: $t = 3.447$, $p = 0.0036$; see Table 5-6). Despite this limitation, we believe these preliminary results are intriguing since all three polymorphisms occur on the same linkage group believed to contain a *cis*-regulatory element that modulates cichlid opsin expression (LG 5), and all three are associated with opsins whose expression is significantly associated with this factor in the hybrid cross (Carleton et al. 2010). Additionally, *SWS2A*-217 obliterates a CRX binding site in numerous cichlids, and polymorphisms affecting CRX binding sites have been shown to modulate *SWS2* opsin expression in zebrafish (Takechi et al. 2008). CNE 10-570 is also found very near the LWS LCR and could also act to affect LCR binding. It is therefore possible that all three alleles acts as, or are linked to, *cis*-regulatory elements that modulate opsin expression in cichlids. In summary, we found some evidence of binding site turnover of CRX binding sites within the 5' promoters of Lake Malawi cichlids, but no evidence of turnover in other candidates TFBS or miRNA target sites. Additionally, we also identified three putative *cis*-regulatory alleles that modulate *SWS2A*, *RH2B*, and *LWS* opsin expression. Although preliminary, these results offer intriguing candidates for additional functional and association analyses in a larger sample of species and individuals.

We acknowledge that the sample sizes we use for phylogenetic shadowing among Lake Malawi's cichlids are small and at best provide a weak test for *cis*-regulatory alleles associated with opsin expression. Additionally, we use cross

species and genera comparisons for an analysis that is generally based around individuals and populations. However, Lake Malawi cichlids are extremely similar at the genetic level and share many ancestral polymorphisms (Loh et al. 2008). For this reason, genetic analyses across cichlid species and genera are analogous to within-species polymorphism studies in other vertebrates, including chimps and humans (Loh et al. 2008; Loh et al. 2010). Additionally, recent work in cichlids has successfully used cross-species comparisons to fine-map *cis*-regulatory alleles underlying pigmentation differences, so long as these differences have a common origin among the different species sampled (Roberts et al. 2009). It is hard to predict which traits will have a common origin among African cichlids, as previous work (Allender et al. 2003) suggested that the pigmentation trait mapped in Roberts et al. (2009) had evolved several times. Our recent work reconstructing the evolution of opsin regulatory changes in cichlids revealed that the three opsin expression palettes have evolved repeatedly among cichlids in Lakes Tanganyika and Malawi (O'Quin et al. 2010), but it is still unclear whether or not the three palettes have a common origin among Lake Malawi's cichlids. But despite our small sample size, we have identified several promising polymorphisms located directly upstream of the opsins that are intriguing candidates for additional analyses. The multiple cichlid genome assemblies and transcriptome analyses that are forthcoming from the Broad Institute will soon allow us to explore regulatory regions farther afield from those studied here in a much larger panel of cichlid individuals and species.

The Search For Cis-*regulatory Sequences*

Cis-regulatory sequences may reside many kilobases away from the genes

they regulate, as in the case of enhancer or repressor elements; or they may be found very near their genetic targets, as in the case of promoter elements and UTRs. So, given this diversity, is it possible to predict which non-coding regions are most likely to contain functional *cis*-regulatory alleles? If we accept estimates of pairwise sequence divergence (D_{xy}) as indicative of those regions most likely to contain functional opsin regulatory alleles, then our estimates of D_{xy} between *O. niloticus* and *M. zebra* suggest that the proximal promoter regions are most likely to contain *cis*-regulatory alleles that alter opsin expression (Figure 5-5A; see also Table S5-6 [Appendix 4] for a list of D_{xy} values for every region examined). The proximal promoters exhibit the highest levels of pairwise sequence divergence of all coding and non-coding regions examined, and also contain more sequences with divergent TFBS profiles (Figure 5-3; Table 5-3), and putative regulatory alleles (Table 5-6). However, this conclusion is undoubtedly influenced by what could be a naive choice of promoter sequences (the true functional opsin promoter regions have not yet been identified in cichlids and may be more highly conserved), increased length of the promoter sequence relative to other regions analyzed (we analyzed 1 kb for each promoter versus ~ 400 bp for each CNE and UTR), and the increased power to detect significant divergence from null expectations afforded by the large number of TFBS found within the proximal promoters (we found ~ 22 TFBS within each promoter versus ~ 6 TFBS/microRNA target sites within each CNE and UTR).

If the overall proportion of divergent TFBS/microRNA target sites (P_{div}) is used to identify those non-coding regions most likely to contain functional *cis*-regulatory alleles, the proximal promoter regions still exhibit the highest proportion

of divergent regulatory regions, although the advantage is only slight. Only about 55% of TFBS are shared between *O. niloticus* and *M. zebra* promoters, while 45% are divergent (Figure 5-5B). In contrast, the CNEs and 3'-UTRs exhibit lower (and very similar) proportions of shared versus divergent TFBS/microRNA target sites (~ 67% shared and ~ 33% divergent; Figure 5-5B). In this case, it is difficult to confidently conclude that 5' promoter regions are more likely to contain functional alleles that regulate opsin expression, although the data are suggestive. When both pairwise divergence and the proportion divergent TFBS/microRNA target sites are taken into account, we find that regions that exhibit statistically significant divergence are not necessarily those regions that exhibit greater pairwise sequence divergence (Figure 5-5C). In fact, the regions with the highest P_{div} also exhibit some of the lowest D_{xy} values. This result suggest that the increased number of statistically divergent promoter regions we observe is not a function of sequence divergence, but rather increased statistical power afforded by the greater length of the sequences surveyed and the greater number of TFBS found.

Additionally, our results show that the majority of the non-coding regions examined exhibit P_{div} values near 37%, with a median of 30% (Figure 5-5C). This observation suggests that the 8% divergence criterion we used as null model for evolutionary divergence is likely too low and also suggests that our power for many regions was lacking due to the small number of TFBS or microRNA target sites identified (see above). But even when a more liberal null divergence value of 30% is used, our results largely remain consistent: CNEs 3 and 6 (located near *SWS1* and *SWS2A*), the proximal promoters for *RH2B* and *SWS1*, and the 3'-UTR for *SWS2B*

still exhibit significant divergence in their number of TFBS and microRNA target sites shared between *O. niloticus* and *M. zebra* ($p < 0.05$ for all comparisons).

Finally, we note that many putative regulatory regions identified in our opsin-containing BACS are highly conserved among many phenotypically diverse cichlid species from Lake Malawi, as well as between the ~ 18 MY divergent *Oreochromis niloticus* and *Metriaclima* zebra. This conservation suggests that *trans*-acting factors may also play an important role in generating evolutionary changes in cichlid opsin expression. For example, in both yeast and humans, interspecific differences in gene expression are primarily the result of *trans*-regulatory factors ((Morley et al. 2004; Sung et al. 2009). And although *cis*-regulatory alleles contribute more to interspecific differences in gene expression among several *Drosophila* species, *trans*-acting alleles generally contribute to these differences as well (Wittkopp et al. 2004). Coding mutations within *trans*-acting transcription factors can act in a modular fashion, thereby mitigating negative pleiotropic effects (Hsia and McGinnis 2003), and these mutations may still affect gene expression even when the sites they bind remain conserved (Levine and Tjian 2003), as many of the TFBS we examine are. Also, in addition to the putative *cis*-regulatory factors associated with *SWS2B*, *SWS2A*, and *RH2B* opsin expression in cichlids, Carleton et al. (Carleton et al. 2010) also identified one *trans*-acting locus in the same cross, as well as another *trans*-acting locus in a separate cross. These two loci, located on cichlid linkage groups (LG) 4 and 13, do not occur in linkage with the cichlid opsins and explain a higher portion of the variance in opsin expression than the single *cis*-associated factor on LG 5 (Carleton et al. 2010). Whether these sites represent transcription factors,

microRNAs, or other *trans*-acting binding sites is unknown, but several good candidate genes are located in these regions. Future work will aim to map and characterize these putative *trans*-regulatory regions in a variety of cichlid taxa.

Conclusions

Mutations within *cis*-regulatory regions are intriguing candidates for the adaptive evolution of gene expression (Wray 2007). Here we generated and surveyed non-coding sequences surrounding the opsin gene arrays of two African cichlids, *Oreochromis niloticus* and *Metriaclima zebra*. This study is the first to systematically survey the cichlid opsins for putative *cis*-regulatory sequences, and our results suggest that these regions could potentially contribute to variation in cichlid opsin expression. The results of our study reveal: (*i*) the presence of at least four conserved non-coding elements located up- and downstream of the opsins that may function as regulators of opsin expression, including a retinal microRNA and one known opsin enhancer (LWS-LCR), (*ii*) significant divergence and enrichment of transcription factor binding sites within the proximal promoter of several opsins, including many that are differentially expressed among African cichlids, (*iii*) numerous target sites for retinal and sensory organ-specific microRNAs within the 3'-UTR of each opsin, including two UTRs that are significantly divergent in these target sites, and (*iv*) several putative *cis*-regulatory alleles located within the promoters of the *RH2B*, *SWS2A* and *LWS* opsins. Future work will aim to further characterize these candidate *cis*-regulatory sequences, as well as to identify candidate *trans*-acting alleles. Given that spectral sensitivity and opsin expression in

vertebrates can be influenced by coding mutations (Spady et al. 2005; Terai et al. 2006; Hofmann et al. 2009), *trans*- regulatory mutations (Carleton et al. 2010), *cis*-regulatory mutations (Takechi et al. 2008), and possibly microRNAs as well, cichlids may represent an ideal system in which to examine how these various molecular mechanisms interact to influence the evolution of visual system diversity in vertebrates.

Tables

Table 5-1. List of candidate transcription factors surveyed in this study.

Transcription Factor	Symbol	OMIM[1] #	TESS[2] # (mice)	Opsin(s) affected	Ref(s)[3]
Activator Protein 1	AP-1	165160	T00032	*SWS1*	a
Con-rod homeobox-protein	CRX/OTX	602225	T03461	*SWS2*	b
Nuclear Factor kappa B	NFκB	164011	T00588	*SWS1*	a
Photoreceptor-specific nuclear receptor	PNR	604485	T03723[4]	SWS	c
Retinoic Acid Receptor α	RARα	180240	T01327	*SWS1*	d
Retinoic Acid Receptor β	RARβ	180220	T01328	*SWS1*	d
Retinoic Acid Receptor γ	RARγ	180190	T01329	*SWS1*	d
Retinoid X Receptor α	RXRα	180245	T01331	-	-
Retinoid X Receptor β	RXRβ	180246	T01332	-	-
Retinoid X Receptor γ	RXRγ	180247	T01333	SWS	e
Thyroid Hormone Receptor α	THRα	190120	T01173	*SWS1*	f
Thyroid Hormone Receptor β	THRβ	190160	T00851[4]	*SWS1*, *RH2*	f, g

[1] Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/omim)
[2] Transcription Element Search System (http://www.cbil.upenn.edu/cgi-bin/tess/tess)
[3] Reference key:
a: Dann et al. (2004)
b: Takechi et al. (2008)
c: Peng et al. (2005)
d: Browman and Hawryshyn (1994a)
e: Roberts et al. (2005)
f: Browman and Hawryshyn (1994b)
g: Ng et al. (2001)
[4] TESS # for human sequences

Table 5-2. Assembly statistics for *O. niloticus* and *M. zebra* opsin-containing BACs.

Species	Opsin Array	Clone ID	Estimated clone size (bp)[1]	Sequencing method	Contig size (bp)	Reads assembled (%)	GenBank acc. no.
O. niloticus	*SWS1*	T4057DH09	210,000	ABI, 454	171,942	100	JF262087
	SWS2-LWS	T4075AE05	184,000	ABI	171,766	85.1	JF262088
	RH2A-RH2B	T4024BG04	200,000	ABI	185,168	84.2	JF262086
M. zebra	*SWS1*	Mz042C6	87,000	454	77,652	-	JF262085
	SWS2-LWS	Mz045P9	96,000	454	107,685	-	JF262084
	RH2A-RH2B	Mz088M22	133,000	454	48,023, 29,393	-	JF262089

[1] Clone size estimated based on Pulsed Field Gel Electrophoresis

Table 5-3. Comparison of sequence similarity and TFBS/miRNA target site divergence for three putative *cis*-regulatory regions surrounding the opsin arrays of *O. niloticus* and *M. zebra*.

Region		Identity (%)	D_{xy}[1] (%)	Length *On* (bp)	Length *Mz* (bp)	TFBS Divt.	TFBS Shrd	Est. P_{div}[2] (%)	p-value[3]
CNE[4]	1	96.84	3.23	158	158	0	2	0.0	> 0.05
	2	96.06	4.05	330	331	5	9	35.7	> 0.05
	3	96.16	3.94	815	760	9	6	60.0	0.00001
	4	83.46	14.87	132	127	1	1	50.0	> 0.05
	5	96.22	3.88	240	239	2	6	25.0	> 0.05
	6	94.74	4.53	349	359	7	1	87.5	$3.1e^{-6}$
	7	98.31	1.70	240	241	2	0	100.0	> 0.05
	8	96.14	3.97	207	207	1	0	100.0	> 0.05
	9	-	-	300	-	-	-	-	-
	10	97.16	2.89	882	885	1	8	11.1	> 0.05
	11	88.46	4.86	779	799	3	9	25.0	> 0.05
	12	93.93	6.33	313	313	1	3	25.0	> 0.05
	13	97.64	2.40	127	127	0	0	-	-
	14	95.97	4.14	124	124	1	1	50.0	> 0.05
	15	95.53	4.61	246	249	1	3	25.0	> 0.05
	16	97.66	2.37	214	214	1	9	10.0	> 0.05
	17	88.97	4.71	999	1404	1	9	10.0	> 0.05
	18	95.32	4.84	428	428	3	6	33.3	> 0.05
	19	91.21	9.35	182	191	0	2	0.0	> 0.05
	20	96.14	3.96	311	313	2	3	40.0	> 0.05
	21	93.25	7.07	1087	976	5	13	27.8	> 0.05
	22	-	-	69	-	-	-	-	-
	23	98.88	1.13	358	38	1	13	7.1	> 0.05
Proximal Promoter[5]	*LWS*	97.56	2.48	1000	1000	1	16	5.9	> 0.05
	RH2Aα	94.80	5.38	1000	1000	10	11	47.6	0.00001
	RH2Aβ	91.77	8.60	1000	1000	14	19	42.4	$5.8e^{-7}$
	RH2B	61.35	9.40	1000	1000	14	7	66.7	$2.1e^{-10}$
	SWS1	71.49	26.37	1000	1000	15	10	60.0	$3.7e^{-10}$
	SWS2A	97.19	2.87	1000	1000	10	12	45.5	0.00002
	SWS2B	81.96	16.31	1000	1000	4	10	28.6	> 0.05

Continued

Region		Identity (%)	D_{xy}[1] (%)	Length *On* (bp)	Length *Mz* (bp)	TFBS Divt.	TFBS Shrd	Est. P_{div}[2] (%)	p-value[3]
3'-UTR[6]	*LWS*	93.39	6.92	189	189	1	4	20.0	> 0.05
	RH2Aα	94.04	6.21	438	442	4	9	30.8	> 0.05
	RH2Aβ	93.26	7.06	465	460	4	11	26.7	> 0.05
	RH2B	93.15	7.18	310	319	4	4	50.0	0.01542
	SWS1	96.74	3.33	217	242	1	3	25.0	> 0.05
	SWS2A	95.90	4.21	124	137	0	1	0.0	> 0.05

[1] Pairwise sequence divergence between *O. niloticus* and *M. zebra*, corrected for multiple hits.
[2] Actual proportion of divergent TFBSs observed for *O. niloticus* and *M. zebra.*
[3] Bonferroni-corrected p-values for the Binomial Exact Test.
[4] See Table S5-4 [Appendix 4] for individual counts of each TFBS identified for the CNEs.
[5] See Figure 5-3 for individual counts of each TFBS indentified for the proximal promoters.
[6] See Table 5-5 for individual counts of each microRNA target site identified for the 3'-UTRs.

Table 5-4. Conserved microRNA target sites within the 3'-UTRs of each opsin in *O. niloticus* and *M. zebra*.

Opsin	miRNA	Target	Conserved[1]	Function and expression	Ref(s)[2]
SWS1	miR-725	TGACTGAG	GA	Expressed in fins	a
SWS2B	miR-217	ATGCAGTA	GA	Alters *PTEN* exp.; found in eye	b, c
SWS2A	miR-181a	AGAATGTA	DR	T-cell regulation; found in eye	b, d
RH2B	miR-23b	TATGTGAA	TR	Ganglion apoptosis; found in eye	e, f
RH2Aα/β	miR-96	TTGCCAAA	OL	Sensory organ specific; found in eye	g, h
	miR-182a	TTGCCAAA	OL	Sensory organ specific; found in eye	g, h
LWS	miR-728	TTTAGTAA	GA,TN,TR	Unknown; found in eye	a
	miR-722*	GCAAAAAA	TR	Unknown; found in eye	a

[1] Other fish species in which this target site is also found: GA: *G. aculeatus* (stickleback); DR: *D. rerio* (zebrafish); TR: *T. rubripes* (tetraodon); TN: *T. nigroviridis* (pufferfish); OL: *O. latipes* (medaka).

[2] References:
a: Kloosterman et al. (2006)
b: Wienholds et al. (2005)
c: Kato et al. (2009)
d: Li et al. (2007)
e: Guerin et al. (2006)
f: Hackler et al. (2010)
g: Xu et al. (2007)
h: Karali et al. (2007)

[3] This site present in *O. niloticus* only

Table 5-5. Polymorphism statistics for 8 candidate *cis*-regulatory regions in 18 Lake Malawi cichlid species.

Opsin	Length (bp)	S[1]	s[2]	H[3]	π[4]	C[5]	T_D[6]	CRX[7]
SWS1	1000	16	5	17	0.0020	0.983	-1.4424	1
SWS2B	694	2	1	3	0.0006	0.997	0.3137	0
SWS2A	1000	7	1	6	0.0010	0.992	-1.1518	2
RH2B	950	17	3	15	0.0022	0.982	-1.1050	2
LWS	956	12	2	11	0.0012	0.987	-1.2394	0
CNE 10	882	12	1	10	0.0021	0.986	-0.2311	0
SWS2B UTR	442	2	1	3	0.0010	0.995	-0.2183	NA
LWS UTR[8]	436	1	0	2	0.0003	0.998	-0.7139	NA

[1] Total number of segregating sites
[2] Total number of segregating sites that are singletons
[3] Total number of haplotypes
[4] Nucleotide diversity
[5] Sequence conservation
[6] Tajima's D
[7] Total number of segregating sites that interrupt predicted CRX binding sites
[8] Statistics presented for in/del polymorphism

Table 5-6. Results of allelic associated between SNPs underlying peaks of nuclotide diversity and opsin expression in 18 Lake Malawi cichlid species.

Polymorphism distance from TSS	Type	MAF[1]	HWE[2]	r^2	t-value	P-value
SWS1 -54	C→T	0.222	N	-0.279	-0.911	> 0.05
SWS2B -208	C→T	0.441	N	-0.001	-0.064	> 0.05
SWS2B -55	1 bp indel	0.471	N	0.175	0.711	> 0.05
SWS2A -224*	C→T	0.222	N	0.127	1.037	> 0.05
SWS2A -217*	8 bp indel	0.194	N	0.392	1.841	0.087
RH2B -308	C→G	0.167	N	-0.245	-0.893	> 0.05
RH2B -161	C→T	0.111	Y	0.263	3.447	0.004
LWS -208	C→T	0.167	N	0.355	1.002	> 0.05
CNE-10 183	A→T	0.222	N	0.055	-0.673	> 0.05
CNE-10 570	C→T	0.417	N	0.608	2.237	0.041
SWS2B-UTR 197	A→C	0.250	N	0.070	0.430	> 0.05

* These polymorphisms interrupt CRX transcription factor binding sites
[1] Minor allele frequency
[2] Hardy Weinberg Equilibrium

Figures

Figure 5-1. Pairwise comparison of sequence conservation between *O. niloticus* opsin-containing BAC regions and four fish genomes. (A) *SWS1* opsin-containing region. (B) *SWS2-LWS* opsin-containing region. (C) *RH2* opsin-containing region. Top line represents *O. niloticus* BAC sequence. Conserved non-coding elements (CNEs) are numbered and highlighted in red; repetitive sequences are highlighted in green; promoter sequences later examined for interspecific polymorphism are highlighted in blue.

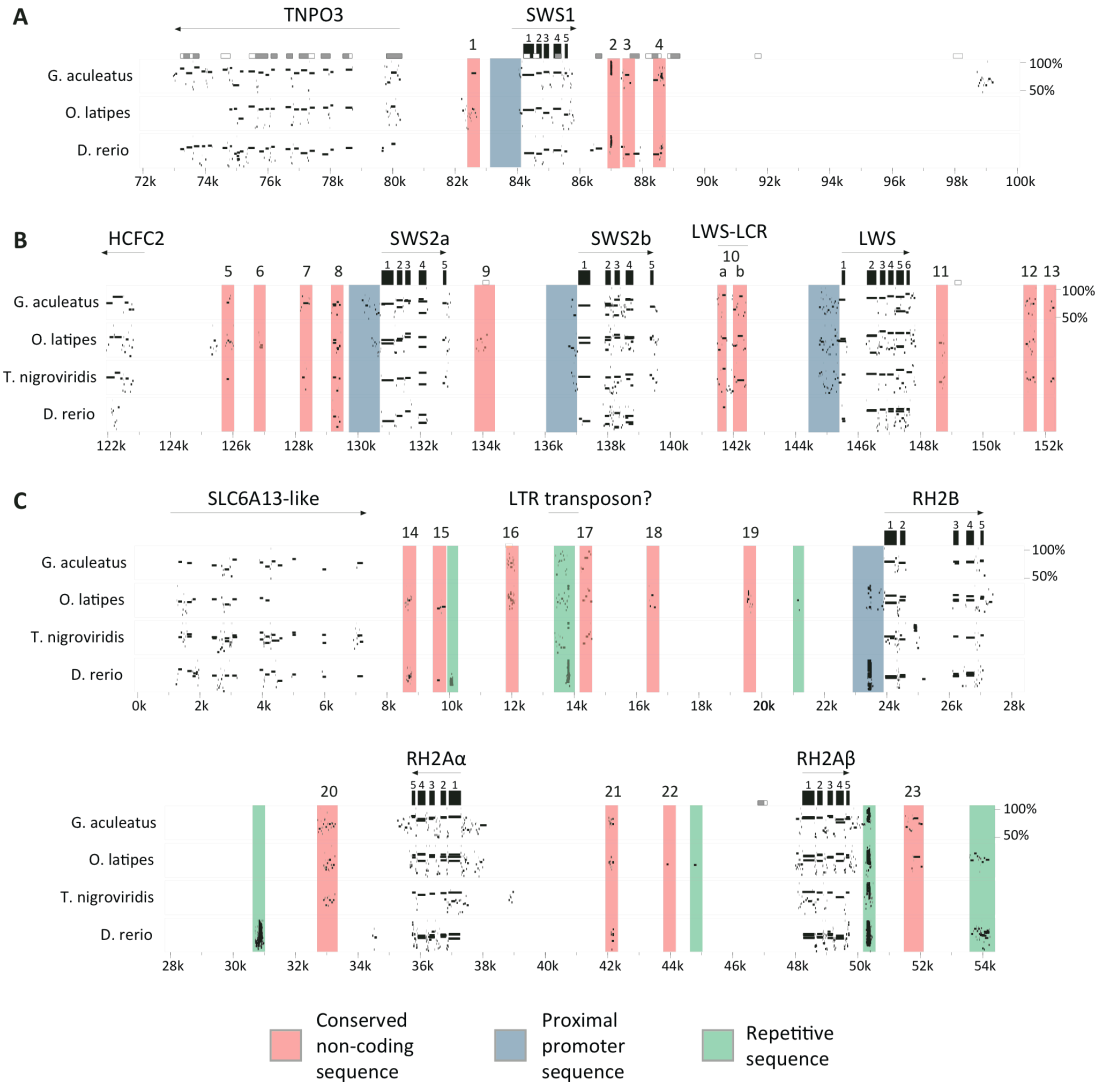


Figure 5-2. Alignment of two putative opsin regulatory elements (CNE 10a and b) in fishes. (A) Alignment of CNE 10a from five fish genomes to dre-miR-726. This region is highly similar among all fish species. Black bar indicates the mature microRNA sequence. (B) Alignment of CNE 10b from five fish genomes to the human LWS-LCR. These sequences show regions of high similarity among humans and fishes. Asterisks (*) indicate positions that are identical among all taxa; colons (:) indicate positions that are identical among the reference and four out of five taxa.

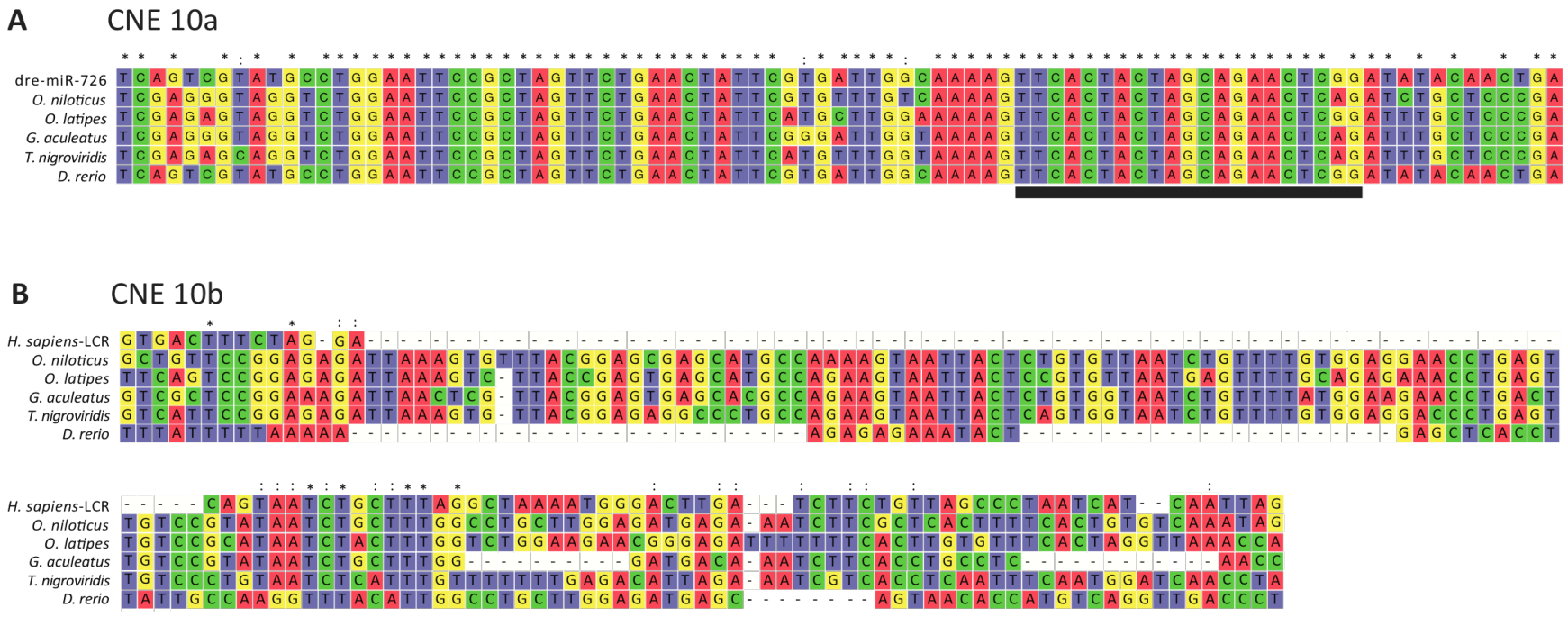


Figure 5-3. Transcription factor binding site diversity within opsin proximal promoters. (A – G) Distribution of ten transcription factor binding sites (TFBS) in the proximal promoters of each opsin in *O. niloticus* and *M. zebra*. TFBS labeled in blue are present in *O. niloticus* only, those in red are present in *M. zebra* only, and those in black are found in both species. Sites labeled simply RAR correspond to all three retinoic acid paralogs. The orientation of factors above or below the central reference line has no special meaning, although *O. niloticus*-only sites are generally above the line, and *M. zebra*-only sites are below it. (H) Comparison of the average number of binding sites for each transcription factor in the proximal promoters of the opsins and seven randomly-selected, non-opsin genes in *O. niloticus*. On average, the opsins contain significantly greater numbers of binding sites for these transcription factors compared to the non-opsin genes.

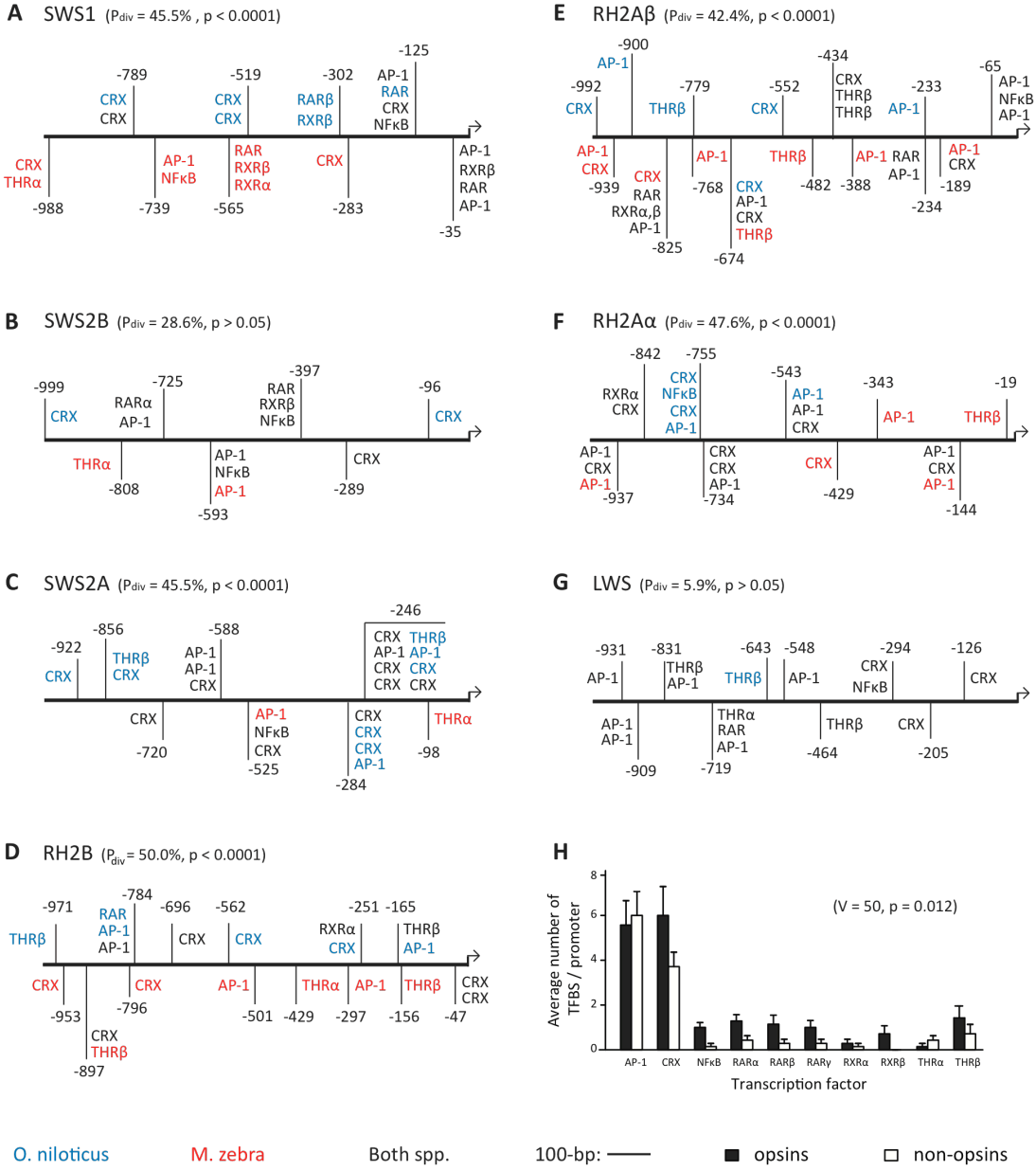


Figure 5-4. Interspecific polymorphism in eight putative *cis*-regulatory regions from 18 Lake Malawi cichlid species. (A – H) Minor allele frequency (MAF; in red) and nuclotide diversity (π; in black) calculated in a sliding window across the proximal promoter regions of five opsins (A – E), CNE 10 (F), and two opsin 3'-UTRs (G – H) using 18 Lake Malawi cichlid species. Numbers above peaks of MAF and π denote the position of SNPs analyzed for allelic-associated with opsin expression (see Table 5-6); asterisks (*) denote polymorphisms that interrupt CRX transcription factor binding sites.

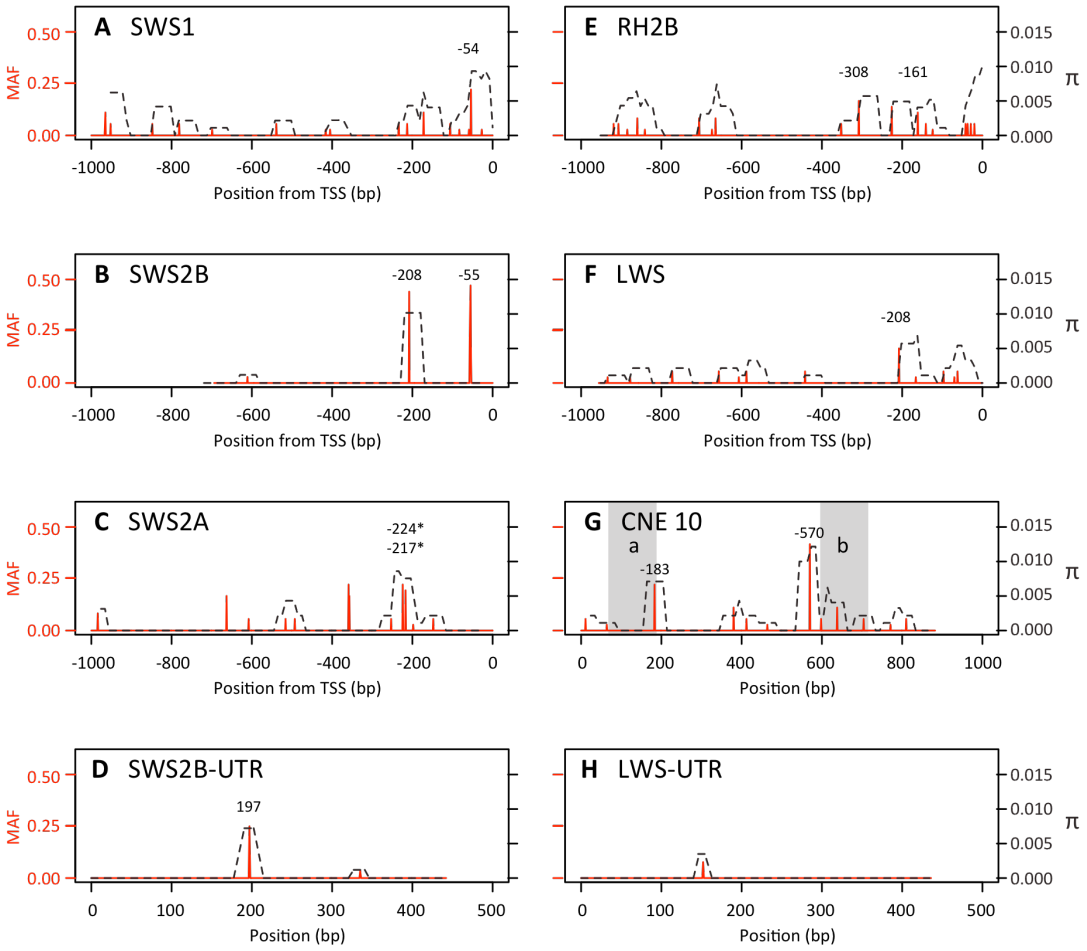
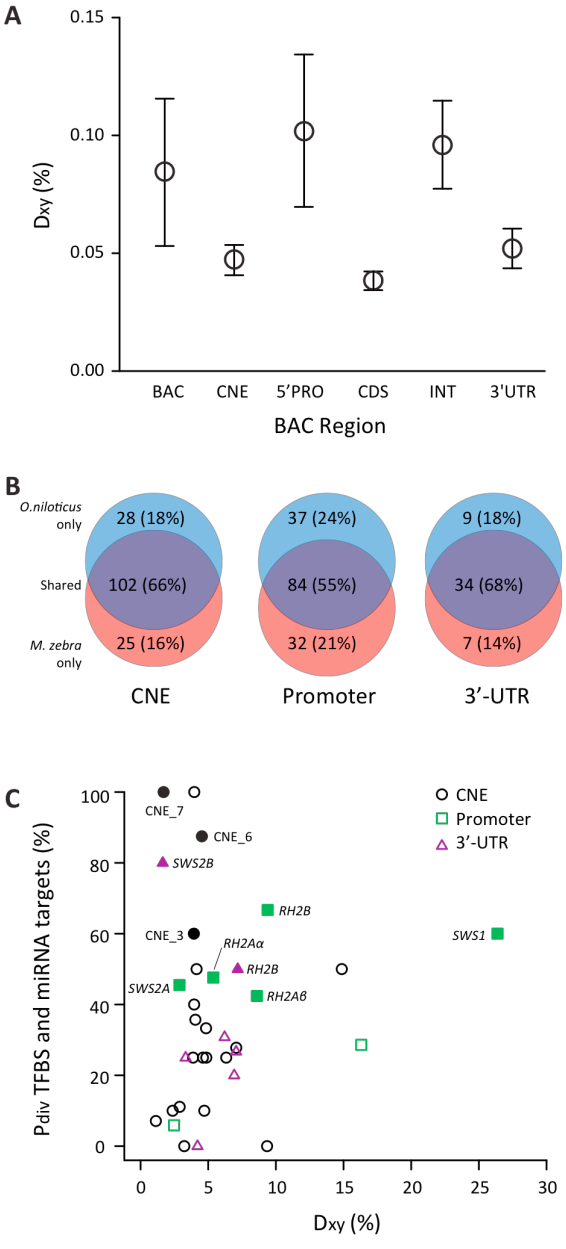


Figure 5-5. Divergence among coding and non-coding regions in *O. niloticus* and *M. zebra* opsin-containing BAC sequences. (A) Pairwise sequence divergence (D_{xy}) between *O. niloticus* and *M. zebra* for different coding and non-coding regions of the opsin-containing BACs. Average Jukes-Cantor-corrected D_{xy} is higher among 5' proximal promoter regions for each opsin. (B) Venn diagram of proportion of shared and divergent TFBS and microRNA target sites among non-coding regions examined in this study. Opsin promoter regions exhibit slightly elevated proportions of divergent sites compared to either CNEs or 3'-UTRs. (C) Comparison of proportion divergent TFBS/miRNA target sites (P_{div}) and pairwise sequence divergence (D_{xy}). Non-coding sequences with elevated P_{div} do not necessarily exhibit increased D_{xy}, even among proximal promoter regions. Filled points are those sequences with P_{div} values that differ significantly from 8% (see Table 5-3).



Chapter 6:

Quantitative Genetic Analysis of *trans*-Regulatory Factors Associated with Opsin Gene Expression in African Cichlids

Kelly E. O'Quin, Zil Patel, Nadia Kahn*, Zan Naseer*, Jane Schulte*, Helena Wang*, Karen L. Carleton

*These authors contributed equally to this work.

Abstract

Recently, several studies have highlighted the prominent role that regulatory mutations can play in phenotypic evolution. However, this body of work has focused primarily on *cis*-regulatory factors, since these mutations do not suffer from mosaic pleiotropy, which presumably affects mutations within *trans*-regulatory transcription factors. Here, we examine what role divergence in 15 candidate *trans*-regulatory factors may play in phenotypic divergence among rapidly radiating African cichlids. Using a hybrid cross of two Lake Malawi cichlid species that differ in the expression of multiple cone opsin genes—the genes responsible for color vision—we show that *trans*-regulatory divergence can play a significant role in phenotypic evolution. Specifically, we find that genotypic variation linked to the transcription factors RORβ and RARγ-2 is statistically associated with variation in the expression of the *SWS2A*, *RH2A*, and *LWS* opsins. Mutations in RORβ have previously been associated with *SWS2A* opsin expression in other systems, but mutations with RARγ-2 have not (although mutations within paralogs of this gene have). Additional work is necessary to confirm that these two transcription factors actually contain mutations causative for variation in cichlid opsin expression; however, we posit that the association of RARγ-2 with opsin expression is evidence that gene duplication can provide a solution to the problem of mosaic pleiotropy for transcription factors. In summary, we find that *trans*-regulatory divergence can play an important role in phenotypic evolution among African cichlids, and likely contributes to phenotypic divergence in other systems as well.

Introduction

Phenotypic evolution can occur through mutations that alter the structure and function of protein-coding genes, or through regulatory mutations that alter their expression (Carroll 2005; Hoekstra and Coyne 2007; Wray 2007; Carroll 2008). The advent of comparative genomics has led to an explosion of studies that address regulatory evolution and the role that these mutations may play in phenotypic divergence. For several reasons, this field has largely focused on the study of *cis*-regulatory mutations (Wray 2007; Carroll 2008). *Cis*-regulatory mutations are found in non-coding DNA, often in tight linkage and close proximity to the genes they regulate, and function by interrupting the binding sites of transcription factors necessary for gene expression (Wray 2007). In contrast, the importance of regulatory mutations within transcription factors has been de-emphasized since many of the factors so far studied (e.g., *Hox* genes) are expressed in many tissues and developmental stages. The broad expression of such genes means that any mutations within them will potentially affect numerous aspects of organismal morphology and physiology, a phenomenon termed "mosaic pleiotropy" (Stern and Tokunaga 1968; Carroll 2008). The consequences for organismal fitness due to mutations in mosaically pleiotropic genes can be dire. But *cis*-regulatory mutations limit the negative consequences of mosaic pleiotropy by altering the expression and function of genes in only a small number of tissues and contexts (Stern 2006; Wray 2007; Carroll 2008). Combined with the observation that *cis*-regulatory mutations are also typically co-dominant, making them an efficient target for natural selection, some authors have argued that *cis*-regulatory mutations are perhaps the most important class of mutation for generating change in animal form (Britten and Davidson 1971; Carroll

2005). Other authors, however, argue that this conclusion is premature, noting that regulatory mutations within transcription factors can also be modular (Hoekstra and Coyne 2007; Oakley 2007). Regardless, the contribution that *trans*-regulatory mutations can make to phenotypic evolution remains poorly understood and potentially under appreciated.

Here we examine what role several transcription factors play in the evolution of opsin gene regulation among African cichlid fishes. African cichlids are a group of freshwater teleost fish found throughout the lakes and rivers of Africa (Fryer and Iles 1972). Within the last 10 million years, cichlids in three African Great Lakes—Lakes Tanganyika, Malawi, and Victoria—have undergone a dramatic adaptive radiation in behavior, color, and form (Kocher 2004; Sturmbauer 2005; Seehausen 2006). But in addition to variation in traits such as body and jaw shape (Albertson et al. 2003), male pigmentation (Allender et al. 2003), and breeding behavior (Kuwamura 1986), African cichlids also exhibit dramatic variation in their color vision (Carleton et al. 2006; Jordan et al. 2006; Carleton 2009; Hofmann et al. 2009). This drastic variation in color sensitivity is largely mediated by variation in the expression of seven cone opsin genes, the genes responsible for absorbing and responding to light.

Cichlids have seven opsin genes used for color vision, each sensitive to a different wavelength or color of light (Spady et al. 2006). These cichlid opsins are: *SWS1* (ultraviolet), *SWS2B* (violet), *SWS2B* (blue), *RH2B* (blue-green), *RH2A*α and *RH2A*β (green), and *LWS* (red) (Carleton and Kocher 2001; Spady et al. 2006; Hofmann et al. 2009; O'Quin et al. 2010). Additionally, cichlids also have an eighth opsin, *RH1*, used for dim-light (non-color) vision. The opsins are expressed within the photoreceptors

cells of the retina. Cichlids have two types of photoreceptors used for color vision. The *SWS1*, *SWS2B*, and *SWS2A* opsins are expressed in single-cone photoreceptor cells, while *RH2B*, *RH2A*, and *LWS* opsins are expressed in double-cone photoreceptor cells (Fernald and Liebman 1980; Fernald 1981; Jordan et al. 2006; Carleton 2009). A recent preliminary survey of quantitative trait loci (QTL) controlling opsin expression in cichlids revealed at least one potentially *cis*-acting QTL linked to the opsin genes *SWS2A*, *SWS2B*, and *LWS* on cichlid linkage group (LG) 5, and one *trans*-acting QTL on LG 13 (Carleton et al. 2010). A follow-up study of putative *cis*-regulatory sequences surrounding the opsins identified multiple divergent sequences and three putative *cis*-regulatory alleles near the *RH2*, *SWS2A*, and *LWS* opsins (both on LG 5) [see Chapter 5]. Here, we perform a reverse QTL analysis (see Mackay and Langley (1990)) in which we expand these earlier studies by asking whether genetic variation linked to several candidate transcription factors is associated with additional *trans*-acting QTL for cichlid opsin expression.

Opsin expression in vertebrates is under considerable *trans*-regulatory control mediated by complex gene regulatory networks (Swaroop et al. 2010). Of the four vertebrate opsin classes (*SWS*, *RH2*, *LWS*, and *RH1*), *trans*-regulatory factors have been shown to control all of them, though *cis*-regulatory factors also play a role (Wang et al. 1992; Tsujimura et al. 2007; Takechi et al. 2008; Tsujimura et al. 2010). The most comprehensive example is the mouse system, where researchers have demonstrated that *RH1* expression is influenced by the transcription factors CRX, NRL, RORβ, and NR2E3 (Furukawa et al. 1997; Mears et al. 2001; Peng et al. 2005; Jia et al. 2009); *SWS* expression is influenced by RORβ, CRX, RXRγ, and RORα (Roberts et al. 2005;

Srinivas et al. 2006; Fujieda et al. 2009); and *RH2/LWS* expression is influenced by RORα and THRβ (Ng et al. 2001; Fujieda et al. 2009) (see also Swaroop et al. (2010) for an overview of transcription regulation of mouse opsins). In some cases, the same or paralogous factors are also known to influence opsin expression in other systems, including humans and zebrafish (Milam et al. 2002; Takechi et al. 2008). Finally many of these genes, as well as their paralogs, are also known to influence more general aspects of vertebrate eye development and patterning (Kastner et al. 1994; Tini et al. 1995; Hyatt and Dowling 1997). Thus, these genes and their paralogs serve as good candidates for the *trans*-regulatory control of opsin expression in cichlids.

To investigate what role these and other transcription factors (see Table 6-1) may play in the *trans*-regulation of cichlid opsin expression, we genotyped genetic markers at or near each gene in a genetic cross of two cichlids with alternate patterns of opsin gene expression (Carleton et al. 2010). *Aulonocara baenschi* is a small, sand-dwelling cichlid from Lake Malawi that is maximally sensitive to violet and green light. *A. baenschi* predominantly expresses the *SWS2B*, *RH2B*, and *RH2A* opsins (Hofmann et al. 2009; Carleton et al. 2010). *Tramitichromis intermedius* is also a small, sand-dwelling cichlid from Lake Malawi, but it is maximally sensitive to blue, green, and red light. *T. intermedius* predominantly expresses the *SWS2A*, *RH2A*, and *LWS* opsins (Parry et al. 2005; Hofmann et al. 2009; Carleton et al. 2010). We used the F_2 intercross progeny of this cross to test whether genotypic variation linked to each individual transcription factor is associated with phenotypic variation in the expression of the seven cichlid opsin genes. Although the results of this analysis cannot definitively determine whether genetic variation in each transcription factor contributes to variation in cichlid opsin expression,

they should shed additional light on what role *trans*-regulatory QTL play in the phenotypic divergence of closely related species.

Materials and Methods

Sampling

We sampled 160 F_2 progeny from a hybrid intercross of two African cichlids that express alternate opsin expression palettes. *Aulonocara baenschi* expressed the middle-wavelength-senesitive opsin set (*SWS2B*, *RH2B*, and *RH2A*) while *Tramitichromis intermedius* expresses the long-wavelength-sensitive opsin set (*SWS2A*, *RH2A*, and *LWS*). A preliminary study of opsin expression among 50 F_2 from this cross was previously used to determine whether *cis*- or *trans*- acting loci contribute to cichlid opsin expression (Carleton et al. 2010). The 160 F_2 progeny chosen for this study were sampled across four F_0 families.

Opsin Gene Expression and Estimated Photoreceptor Sensitivity

We measured opsin gene expression in all 160 F_2 intercross progeny via RT-qPCR following previously published protocols (Spady et al. 2006) [see also Chapters 2 – 3]. Following RT-qPCR, we used the final expression results to estimate the wavelength of maximum absorbance for the single- and double-cone photoreceptors of each individual using the formula:

$$PS_{max,\,C} = \frac{\Sigma(f_i \lambda_i)}{\Sigma f_i}$$

where f_i is the percent expression of the i^{th} opsin in either single- or double-cones (C), and λ_i is the wavelength of maximum absorbance (λ_{max}) of the i^{th} opsin in *O. niloticus* (Spady et al. 2006) [see also Chapters 3 – 4]. We use these estimated photoreceptor sensitivities as a data reduction step, since together they summarize multivariate changes in the expression of six opsins into just two variables.

Marker Selection and Genotyping

To test for association between genetic variation at each candidate gene and estimated single- and double-cone sensitivities, we identified and genotyped a combination of simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) within or near each candidate gene. We started by locating these genes in stickleback (*Gasterosteus aculeatus*) using the UCSC Genome Browser (Broad/gasAcu1 assembly, Feb. 2006; http://genome.ucsc.edu/cgi-bin/hgGateway?db=gasAcu1/) to identify the location of each candidate gene in the stickleback genome (Table 6-1). We then used these stickleback coordinates to search for any cichlid sequence from the same region using an assembly of cichlid (Nile tilapia, *Oreochromis niloticus*) sequences mapped to the stickleback genome (Soler et al. 2010) [see also http://www.bouillabase.org; site accessed June 2010]. We chose assembled cichlid sequences (nodes) that either contained the candidate gene or, if none were available, that were adjacent to the candidate gene in the stickleback assembly. In this latter case, we only chose nodes that were located within ≤ 600 kb of the candidate gene. This distance ensures that any recombination between the marker and candidate gene would be very unlikely (1 cM ≈ 750 kb in cichlids (Lee et al. 2005)), so long as synteny in the region is conserved. Once we identified a cichlid node for a particular candidate gene, we then

searched it for SSR or SNP markers.

To identify SSR markers, we searched each cichlid node for SSRs using the program Perfect Microsatellite Repeat Finder (http://sgdp.iop.kcl.ac.uk/nikammar/repeatfinder.html). We then created PCR primers to amplify a 100 – 400 bp region surrounding the SSR in the program Primer3Plus (Untergasser et al. 2007). We performed all SSR PCR reactions in 20 μl volumes (14.4 μl dH_2O, 2.0 μl 10X Reaction Buffer, 1.6 μl dNTP mix, 0.4 μl each Forward and Reverse primers, 0.2 μl Promega GoTaq™ DNA Polymerase, and 1.0 μl template DNA). All forward primers were fluorescently labeled with 5-HEX or 6-FAM. We initially performed PCR of the SSR on the F_0 individuals only and genotyped the reactions using an ABI 3730 genetic analyzer and GeneMapper® software (Applied Biosystems). If a polymorphism was found in the F_0 individuals, we then moved on to the F_2 intercross progeny.

To identify SNP markers, we again used Primer3Plus to design PCR amplification and sequencing primers to resequence a small portion of the candidate gene (usually one or more introns). We performed all SNP PCR reaction in 25 μl volumes (19 μl dH_2O, 2.5 μl 10X Reaction Buffer, 0.5 μl dNTP mix, 1.0 μl each Forward and Reverse primers, 0.5 μl DyNAzyme™ II DNA polymerase, and 0.5 μl template DNA). Following PCR, we gel-extracted the PCR product from a 1% agarose gel and cleaned the reaction using a QiaQuick Gel-Purification kit (Qiagen). We then performed separate sequencing reactions for each forward and reverse primer using a BigDye® v3.1 Cycle Sequencing kit. We performed all sequencing reactions in 10 μl volumes (2.5 μl dH_20, 1.5 μl 5X sequencing buffer, 1.0 μl primer, 1.0 μl BigDye®, and 4.0 μl template DNA). Following

cycle sequencing, we cleaned each reaction via sodium acetate/EDTA/ethanol precipitation. We sequenced and genotyped these reactions using an ABI 3730 genetic analyzer and Sequencher® software (GeneCodes, Ann Arbor, MI). Table 6-2 lists all primers used to amplify and genotype the SNP and SSR polymorphims used in this study.

Statistical Analysis of Quantitative Trait Loci

We searched for the presence of *trans*-acting QTL at each candidate gene using analysis of variance (ANOVA). This method provides a simple test of association between the phenotypic variation in the expression of each opsin and genetic variation at the genetic markers linked to each candidate gene (Lynch and Walsh 1998). We initially performed two ANOVAs for each polymorphism, one for both estimated single- and double-cone sensitivity. If any one test was moderately significant ($P < 0.15$), we then repeated the analysis on the individual opsin genes expressed in that photoreceptor type (single-cone: *SWS2B*, *SWS2A*; double-cone: *RH2B*, *RH2A*, *LWS*). Following all hypothesis testing, we used a Bonferroni-corrected threshold to determine the statistical significance of the final results. For markers that were significantly associated with opsin expression following this correction, we estimated the dominance coefficient, *k*, of the alleles at each marker using the single-locus estimation equation outlined in Lynch and Walsh (1998, pp. 61-63).

Additionally, for each transcription factor we started our statistical analysis by selectively genotyping 32 F_2 that recapitulate the F_0 phenotypes. If a marker exhibited moderate association in this initial panel ($p < 0.15$), we then genotyped it in the remaining 128 F_2. For this reason, the number of individuals used across markers may vary considerably. Also, we note that nearly all the candidate genes chosen for this study

map to different cichlid linkage groups than the QTL previously identified (LGs 5 and 13; Carleton et al. (2010)), except RARγ-2 and VSX (Table 6-1). VSX is located between two markers associated with opsin expression on LG 13, and RARγ-2 is located between two tandem arrays of opsins on LG 5. One marker located near the *SWS2-LWS* opsin array on LG 5 has previously exhibited association with *SWS2B*, *SWS2A*, and *RH2B* opsin expression (Carleton et al. 2010).

Results

Quantitative Variation in Opsin Expression Among F_2 Hybrids

We found that opsin expression varied considerably among the F_2 individuals sampled, except for the *SWS1* opsin, which was not highly expressed in either the F_2 or F_0 (Figure 6-1). Of the remaining five opsins, the greatest variation was seen in the expression of the three double-cone opsins, *RH2B*, *RH2A*, and *LWS*. Variation in *RH2A* and *LWS* expression varied continuously among the F_2 progeny, while variation in *SWS2B*, *SWS2A*, and *RH2B* expression was generally skewed towards many observations with little or no expression (Figure 6-1). Predicted single- and double-cone sensitivities estimated from these values were continuous and relatively normally distributed among the F_2 individuals (Figure 6-1). Interestingly, in many cases the single- and double-cone sensitivities we estimated for the F_2 fall well outside of any values estimated for > 60 wild-caught cichlid species from Lake Malawi (Hofmann et al. 2009) [see Chapter 2] (Figure 6-1C inset).

Genotypic Variation at Candidate Gene Marker Loci

We genotyped SSR and SNP polymorphisms for 15 candidate genes in our F_2 intercross hybrids (Table 6-1); in total, we scored genotypes from > 1,300 genotyping reactions. Polymorphisms at most markers were not differentially fixed between the F_0 individuals used. Many loci were heterozygous in one or both species, resulting in loci that were not fully informative. Where possible, we used the genotypes of the parents to attribute one allele to each species (*Aulonocara* or *Tramitichromis*). If an allele was present in both parents, we labeled any offspring with that allele as missing data, since we could not determine whether it was homo- or heterozygous for the parental genotypes. Sorting the alleles of parents in this way reduced some of our statistical power to detect QTL at each marker when the locus was not fully informative (see degrees of freedom in Table 6-3), but was necessary to infer the directional effect of each species' alleles on opsin expression.

QTL Analysis at Candidate Gene Marker Loci

Analysis of variance (ANOVA) of estimated single- and double-cone sensitivities among the genotypic classes at each marker revealed five candidates linked to potential quantitative trait loci (QTL) (Table 6-3). One SSR marker located near RORβ was moderately associated with estimated single-cone sensitivity ($F_{2, 32} = 2.467$, $p = 0.101$). In contrast, four markers exhibited moderate association with estimated double-cone sensitivity. These markers include SSRs near NCOA1, RARβ-2, RARγ-1, and RARγ-2 (all $p < 0.15$) (Table 6-3). Although no single marker achieved association with estimated single- or double-cone sensitivity at $p < 0.05$, we chose to use these five moderately-associated markers as candidates in an additional ANOVA with each of the

six opsins expressed in single- and double-cone photoreceptors.

For the lone marker associated with estimated single-cone sensitivity, we compared mean opsin expression among genotypes for the *SWS2B* and *SWS2A* opsins only, since *SWS1* was not highly expressed in the F_2 progeny. Prior to this analysis, we transformed the opsin expression values by raising *SWS2B* to the 1.308^{th} power and *SWS2A* to the 0.45^{th} power. These transformations, estimated in the R package 'car' (Fox 2008), help meet the ANOVA assumptions of normality and homogeneity of variances. After performing these transformations, we found that mean *SWS2A* expression differed significantly among the genotypic classes at the RORβ locus ($F_{2, 32} = 3.949$, $p = 0.029$), but mean *SWS2B* expression did not ($F_{2, 32} = 2.582$, $p = 0.091$). Genotypic variation at the RORβ marker explained 19.8% of the variation in *SWS2A* expression, revealing the presence of a QTL of moderate affect in the area of this marker. Alleles derived from *Tramitichromis* were associated with higher levels of *SWS2A* expression (Figure 6-2), consistent with observations from the adults of these species (Hofmann et al. 2009). The single-locus estimate of the dominance coefficient, *k*, for these genotypes revealed that the *Tramitichromis*-derived alleles are slightly dominant ($k = 0.327$; see Figure 6-2); however, we note that the phenotypic midpoint we observe for AT individuals at this locus is not significantly different from that estimated for complete additivity (data not shown).

For the four markers associated with estimated double-cone sensitivity, we compared mean opsin expression among genotypes for all three double-cone opsins (*RH2B*, *RH2A*, and *LWS*). Prior to analysis, we transformed *RH2B* expression by raising it to the -7.643^{rd} power, as estimated in the 'car' package; we did not transform *RH2A* or

LWS expression. Once again, this transformation ensured that all ANOVA assumptions regarding the distribution of errors were met. We found that mean opsin expression varied among the genotypes of only one marker, RARγ-2 (Table 6-4). Genotypes at the marker located near RARγ-2 differed significantly in both mean *RH2A* ($F_{2, 150} = 6.57$, $p = 0.002$) and *LWS* ($F_{2, 150} = 5.794$, $p = 0.004$) expression. No candidate gene marker was significantly associated with *RH2B* expression (Table 6-4). Genotypic variation at RARγ-2 explained 8.1% and 7.2% of the variation in *RH2A* and *LWS* opsin expression, respectively, revealing the presence of a minor-effect QTL in the region of this marker. Alleles derived from *Tramitichromis* contributed to an increase in *RH2B* expression, along with a nearly equal decrease in *LWS* expression (Figure 6-2). This pattern is opposite to the direction observed in the parental species, where *T. intermedius* exhibits greater *LWS* expression than *A. baenschi.* The single-locus estimate of *k* revealed that the *Aulonocara*-derived alleles were dominant in both cases ($k = 0.364$ and 0.684, respecitively; see figure 6-2), and that the phenotypic midpoints observed for AT individuals at both loci were significantly different from the midpoints estimated for complete additivity (data not shown). The similarity of the *RH2A* and *LWS* results could be due to a genetic and developmental correlation between these two opsins as observed in other cichlids (Carleton et al. 2008; O'Quin et al. 2011) [see also Chapter 4], or possibly due to the use of a common regulatory switch to simultaneously enhance the expression of one opsin (*RH2A*) while repressing the other (*LWS*).

Discussion

Trans-*acting QTL and the genetic basis of cichlid opsin expression*

We tested 15 candidate *trans*-acting loci for linkage to opsin expression in African cichlids and found at least two *trans*-acting QTL. One QTL is associated with *SWS2A* expression and is linked to the nuclear hormone receptor RORβ. A second QTL is associated with both *RH2A* and *LWS* expression and is linked to the nuclear hormone receptor RARγ-2. These two genes likely reside on cichlid linkage groups (LG) 12 and 5 respectively (Table 6-1). A previous survey of opsin expression in this cross identified two QTL linked to opsin expression, one *trans*-acting locus on LG 13, and one putative *cis*-acting locus on LG 5 near the *SWS2-LWS* opsin array (Carleton et al. 2010). Interestingly, a candidate homeobox gene (VSX) found within the QTL on LG 13 did not exhibit association with opsin expression in this study (Table 6-3), possibly because of an inversion in this region relative to stickleback (Carleton et al. 2010). Taken together, our current and previous studies reveal that the genetic basis of cichlid opsin expression is oligogenic and may be dominated by *trans*-acting QTL.

Both RORβ and RARγ-2 are involved in retinoic acid reception and signaling. Retinoic acid is an important signaling hormone in vertebrates and is known to affect numerous developmental and adult phenotypes (Lammer et al. 1985; Durston et al. 1989; Papalopulu et al. 1991); additionally, retinoic acid has also been shown to regulate the expression of > 500 genes (Balmer and Blomhoff 2002). The nuclear hormone receptors RORβ and RARγ-2 both contain ligand- and DNA-binding domains that allow them to bind promoter elements in response to retinoic acid signaling (Evans 1988).

Importantly, one of these candidate genes, RORβ, has been directly associated

with vertebrate opsin expression in previous studies. RORβ activates *SWS* expression in mice in combination with the cone-rod homeobox gene, CRX (Srinivas et al. 2006). In our study, genotypic variation linked to RORβ is associated with greater *SWSA* expression among the F_2 with *Tramitichromis*-derived alleles (Figure 6-2). Unfortunately, our previous survey of transcription factor binding sites in the opsin promoters of cichlids did not survey the RORβ gene [Chapter 5]. However, this previous study did survey other transcription factors, including thyroid hormone beta (THRβ) and CRX, and found two and five binding sites for these two factors upstream of the *SWS2A* opsin in another Lake Malawi cichlid, *Metriaclima zebra* [see Chapter 5, Figure F-X]. Srinivas et al. (2006) demonstrated that RORβ binds the response element 5'-AGGTCA-3', which is also a core motif of the thyroid hormone response element (Umesono et al. 1991). Thus, it is reasonable to conclude that the THRβ binding sites previously found upstream of the cichlid *SWS2A* opsin also serve as RORβ binding sites. Therefore, it is plausible that RORβ could regulate *SWS2* opsin expression in cichlids.

In contrast, RARγ-2 has not previously been associated with opsin expression in vertebrates, though its close paralogs THRβ and RXRγ have (Roberts et al. 2005). In mice, THRβ and RXRγ suppress SWS opsin expression and promote MWS (*RH2*/*LWS*) opsin expression (Roberts et al. 2005; Swaroop et al. 2010). Fitting with this pattern, we find that genotypic variation linked to RARγ-2 is associated with *RH2* and *LWS* expression (Table 6-4; Figure 6-2). Additionally, our previous analysis of transcription factor binding sites upstream of the cichlid opsins found two RARγ binding sites upstream of the *RH2Aβ* opsin and one upstream of the *LWS* opsin. Thus, it is again plausible that RARγ-2 could regulate opsin expression in cichlids. Although RARγ-2 has

not itself been associated with opsin expression in vertebrates, we note that considerable functional redundancy exists among the various RAR, THR, and RXR gene families and their paralogs. These transcription factors often work synergistically as heterodimers to mediate gene expression in response to retinoic acid signaling (Kastner et al. 1994; Chiba et al. 1997), and both the RARs and RXRs respond to the ligand all-*trans* retinal (Repa et al. 1993).

Mosaic Pleiotropy, Gene Duplication, and Cichlid Vision

Retinoic acid is an important modulator of vertebrate morphogenesis, affecting development of limbs, the central nervous system, and gene expression (Lammer et al. 1985; Durston et al. 1989; Papalopulu et al. 1991); additionally, retinoic acid and is also known to influence photoreceptor development and opsin expression (Dräger and McCaffery 1997; Hyatt and Dowling 1997). Given the important developmental role for retinoic acid and its receptors, it is perhaps unsurprising that the various RAR and RXR genes exhibit mosaic pleiotropy. Mice mutant for multiple RAR or RXR deletions exhibit severe developmental defects in many different tissues and systems, including the eye, skull, lungs, and urogenital tract (Lohnes et al. 1994; Luo et al. 1996). But significantly, mutations in individual RAR or RXR genes often do not produce these results, revealing some level of functional redundancy in these genes (Li et al. 1993; Lufkin et al. 1993; Kastner et al. 1994; Cammas et al. 2010). We feel this observation is significant because one argument often used to support a prominent role for *cis*-regulatory mutations in phenotypic evolution is that, through modularity, they are able to limit the negative consequences of mosaic pleiotropy (Stern 2006; Carroll 2008). In contrast, these same authors argue that mutations within widely-expressed transcription

factors will uniformly change the function of that gene in many developmental stages and tissues, while also possibly altering the expression of hundreds of genes (Carroll 2008). Therefore, mutations in such mosaically pleiotorpic genes are expected to have substantial negative consequences for organismal fitness. One mechanism around this problem, however, is gene duplication (Hoekstra and Coyne 2007). Through duplication, paralogous genes are free to evolve novel functions as long as functionally redundant copies still exist in the genome (Ohno 1970; Lynch and Conery 2000). Therefore, if genotypic variation within RARγ-2 does in fact contribute to phenotypic variation in cichlid opsin expression, then this observation would support the model of phenotypic divergence through the duplication of core developmental genes. Further, we note that the RARs can form multiple isoforms (Giguère et al. 1990; Kastner et al. 1990; Leroy et al. 1991; Giguère et al. 1994). In addition to gene duplication, alternative splicing provides yet another mechanism whereby the negative consequences of mosaic pleiotropy are reduced for functionally conserved proteins (Hoekstra and Coyne 2007).

Future Analyses of Cichlid Opsin Expression with RADseq

However, additional work remains before the divergence of cichlid opsin expression can be confidently attributed to the evolution of transcription factor paralogs. In particular, RARγ-2 may reside on the same linkage group as two of the opsin gene arrays and one previously identified opsin expression QTL. In the medaka genome assembly (*Oryzias latipes*: NIG/UT MEDAKA1/oryLat2 Oct. 25, accessed March 2011), RARγ-2 is located nearly equidistant between the two opsin arrays (*RH2B*-*RH2A* and *SWS2A*-*SWS2B*-*LWS*), suggesting that the same may be true for cichlids and LG 5. But this same region is poorly assembled in the stickleback genome, and the RARγ-2 gene is

placed on an unordered contig. Thus, the position of RARγ-2 in the cichlid genome is currently uncertain. The construction of a larger genetic map of these genes and additional markers should solve this problem, and will help determine whether the QTL linked to RARγ-2 is unique from the QTL previously detected on LG 5.

Additionally, since many of the markers used in this study are merely SSRs found within close genetic proximity to the candidate genes of interest, it is not certain that genetic variation actually exists in many of the genes analyzed, including RORβ and RARγ-2. Therefore, if these genes do harbor causative alleles for cichlid opsin expression, it will be necessary to sequence these in search of nonsynonymous or regulatory mutations that may alter their function in a diverse array of cichlids. Alternatively, these candidates may merely reside in linkage to the true causative alleles which reside within or upstream of other genes. Thus, a larger-scale analysis of genome-wide polymorphisms will be necessary to determine whether these candidate transcription factors truly form peaks of association with cichlid opsin expression. Finally, some additional paralogs of the candidate transcription factors chosen here remain to be tested. Specifically, we left untested the genes PNR-2, RXRα, RXRβ-1 and 2, RORα, VSX-2, and CRX-2 (where the number following each gene denotes a duplicated copy, not an isoform). Thus, additional QTL for cichlid opsin expression may exist at these and other loci, but remain to found.

Fortunately, much of this follow-up work is currently underway. We are sequencing restriction site associated DNA markers (RADseq, see Biard et al. (2008)) in all 160 F_2 intercross progeny used in this study. When complete, we expect to genotype over 25,000 ~100 bp fragments of the cichlid genome, yielding ~2,500 polymorphic

SNPs. The linkage map produced from these markers will be useful for determining the position of RARγ-2 relative to the opsin genes, and the resulting QTL analysis will determine how broad are the peaks of association surrounding each transcription factor. Broader surveys of these markers in all 160 progeny may also increase our statistical power to detect an association between these candidates and the other markers analyzed.

Conclusions

Phenotypic variation may evolve through multiple molecular genetic mechanisms, including mutations to protein-coding genes, *cis*-regulatory sequences, and *trans*-regulatory factors (Hoekstra and Coyne 2007; Wray 2007; Carroll 2008). Although most work on regulatory mutations have focus on *cis*-regulatory alleles, our analysis of opsin expression in a hybrid cross of African cichlids suggest that *trans*-regulatory divergence is also an important source of genetic variation contributing to phenotypic divergence. In particular, in combination with a previous study (Carleton et al. 2010), we find evidence that at least two *trans*-regulatory factors underlie opsin gene expression differences in two cichlid species. In this study, we find evidence that links the transcription factors RORβ and RARγ-2 to *trans*-regulatory variation in cichlid opsin expression. This observation suggests that retinoic acid signaling may be an important factor in determining cichlid opsin expression. Further, the observation that RARγ-2 may be linked to to opsin expression offers preliminary but intriguing evidence that mosaic pleiotropy in *trans*-regulatory factors may be overcome through gene duplication. However, additional work remains before this assertion can be confidently made in the current study. But given that past and present work has shown that cichlid photoreceptor

sensitivity and opsin expression may be altered by mutations within protein-coding genes (Spady et al. 2005; Hofmann et al. 2009; Smith and Carleton 2010), *cis*-regulatory alleles (Carleton et al. 2010) [see also Chapter 5], and also *trans*-regulatory factors (Carleton et al. 2010) [and this Chapter], we believe that the cichlid system may be ideal for dissecting the distinct and shared contributions that these different mutational mechanisms make to phenotypic evolution.

Tables

Table 6-1. Candidate *trans*-regulatory factors examined in Chapter 6.

Candidate	Symbol	Opsin Exp[1]	ENSGACG ID[2]	Cichlid LG
Cone-rod homeobox gene	CRX	*RH1*, *SWS*	00000005793	14
Nuclear receptor coactivator 1	NCOA1	-	00000006727	15
Photoreceptor-specific nuclear receptor	PNR-1	*RH1*	00000017060	1
Retinoic acid receptor alpha	RARα	-	00000005297	4
Retinoic acid receptor beta 1	RARβ-1	-	00000007999	22
Retinoic acid receptor beta 2	RARβ-2	-	00000012955	11
Retinoic acid receptor gamma 1	RARγ-1	-	00000009372	20
Retinoic acid receptor gamma 2	RARγ-2	-	00000000612	5
RAR-related orphan receptor alpha	RORα	*SWS, RH2*, *LWS*	00000010672	7
RAR-related orphan receptor beta	RORβ	*SWS2, RH1*	00000011556	12
Retinoid X receptor gamma	RXRγ	*SWS*	00000011685	23
Thyroid hormone receptor alpha 1	THRα-1	-	00000003766	8
Thyroid hormone receptor alpha 2	THRα-2	-	00000006540	4
Thyroid hormone receptor beta	THRβ	*RH2, LWS*	00000007996	22
Visual system homeobox gene 1	VSX1	-	00000012138	13

[1] These transcription factors have been shown to regulate the expression of opsin genes in various systems, including mice and zebrafish (see Introduction for references)

[2] ID of the stickleback gene in the Ensemble genome browser

Table 6-2. Primers used to amplify and genotype polymorphisms in this study.

Gene	Type	Forward Primer (5'—3')	Reverse Primer (5'—3')
CRX	SSR	TAGCTTAGCAGGGGAGAGCA	CTGGTGGACAAGATGAGCAG
NCOA1	SSR	GCAGCAGAAGCCATGTAGGT	CAACAGGAAACCAACTTTACCAG
PNR-1	SSR	AACACGCAAATCAAGTTCC	TGCCCTTTTTGAACGTTTTT
RARα	SSR	TCATTGCTCTGGATCACACC	TGCGTGACTGGAATGAAGAG
RARβ-1	SSR	GGCTGATGGCCGATATTAAA	TCCCCAGCAACTTTCTTGTT
RARβ-2	SSR	GGAGTCCCAAAACCAGATCA	CGCCTGGATTTTCATTGTTT
RARγ-1	SSR	GAGCCCTGGGTGTTTTAACTT	GAAGGCCAGCATTTTCTTGA
RARγ-2	SSR	GAAGAAGCAACCCACAGAGC	ATCCCTAAACCTCCCACACC
RORα	SSR	GTTGTCCCTGCAAGCTCCTAT	TTGTCACCAGGCACATCATT
RORβ	SSR	TCCATAGAAACACGCACTAACA	TGCAGAGGTTGAAGTGACAAA
RXRγ	SSR	GAGCCCAAATTGTGAGGAAA	TCCAGAACCAAAGAGCCAAA
THRα-1	SSR	GCGCGGAATTCGACGATTCA-GAAGAACCTCCA	GCGCGCAAGCTTCTGAGCAT-TGGTGTGACGAT
THRα-2	SSR	GGCTGAGCACTGTTGCATAA	ACACAGACGAACAGCGTGAT
THRβ	IN/DEL	CCGCATCTGTTGTTTTTTCAT	ATGAAGAACCCGTGTCAAGC
VSX1	SNP	GCGGAGTTGAGGATGGACT	AGCTGGAGGAGCTGGAGAA

Table 6-3. Results of ANOVA comparing mean single- and double-cone sensitivity among genotypic classes at 15 candidate *trans*-regulatory loci.

Candidate	df	Single cone Sensitivity		Double-cone Sensitivity	
		F-value	P-value	F-value	P-value
CRX	2, 23	0.225	0.802	0.342	0.714
NCOA1	1, 148	0.865	0.354	2.822	0.085**
PNR-1	2, 28	1.397	0.264	0.884	0.424
RARα	2, 61	1.378	0.260	1.482	0.235
RARβ-1	2, 107	0.670	0.514	1.467	0.235
RARβ-2	2, 102	1.051	0.353	2.096	0.128**
RARγ-1	2, 143	0.469	0.627	1.952	0.146**
RARγ-2	2, 150	0.153	0.858	2.060	0.131**
RORα	2, 16	1.538	0.245	0.670	0.525
RORβ	2, 32	2.467	0.101**	1.559	0.226
RXRγ	2, 112	0.229	0.796	1.896	0.155
THRα-1	2, 63	0.522	0.596	1.515	0.228
THRα-2	2, 145	0.465	0.629	0.266	0.767
THRβ	2, 65	1.082	0.345	0.967	0.386
VSX1	2, 55	0.417	0661	0.010	0.990

** These results are moderately significant at $p < 0.15$; we followed these results with ANOVA comparing the mean expression of individual opsins (see Results and Table 6-4).

Table 6-4. Results of ANOVA comparing mean *RH2B*, *RH2A*, and *LWS* opsin expression among genotypic classes at four candidate *trans*-regulatory loci.

Candidate	df	*RH2B*		*RH2A*		*LWS*	
		F-value	P-value	F-value	P-value	F-value	P-value
NCOA1	1, 148	0.339	0.561	0.015	0.902	1.544	0.216
RARβ-2	2, 102	0.419	0.659	0.733	0.483	1.883	0.157
RARγ-1	2, 143	0.439	0.651	0.051	0.950	1.213	0.300
RARγ-2	2, 150	0.400	0.671	6.57	0.002**	5.794	0.004**

** These results are significant at the Bonferroni-corrected significance threshold of 0.0042 ($\alpha = 0.05/12$).

Figures

Figure 6-1. Variation in opsin gene expression in F_2 hybrid intercross progeny of a genetic cross used to study cichlid opsin expression. (A) Boxplots of relative gene expression for six opsins. (B) Histograms of single- and double-cone photoreceptor sensitivity estimated from opsin gene expression. (C) Biplot of estimated single- and double-cone sensitivity among F_2 intercross progeny. Variation in single- and double-cone sensitivity observed in >60 wildcaught cichlid species from Lake Malawi is shown inset.

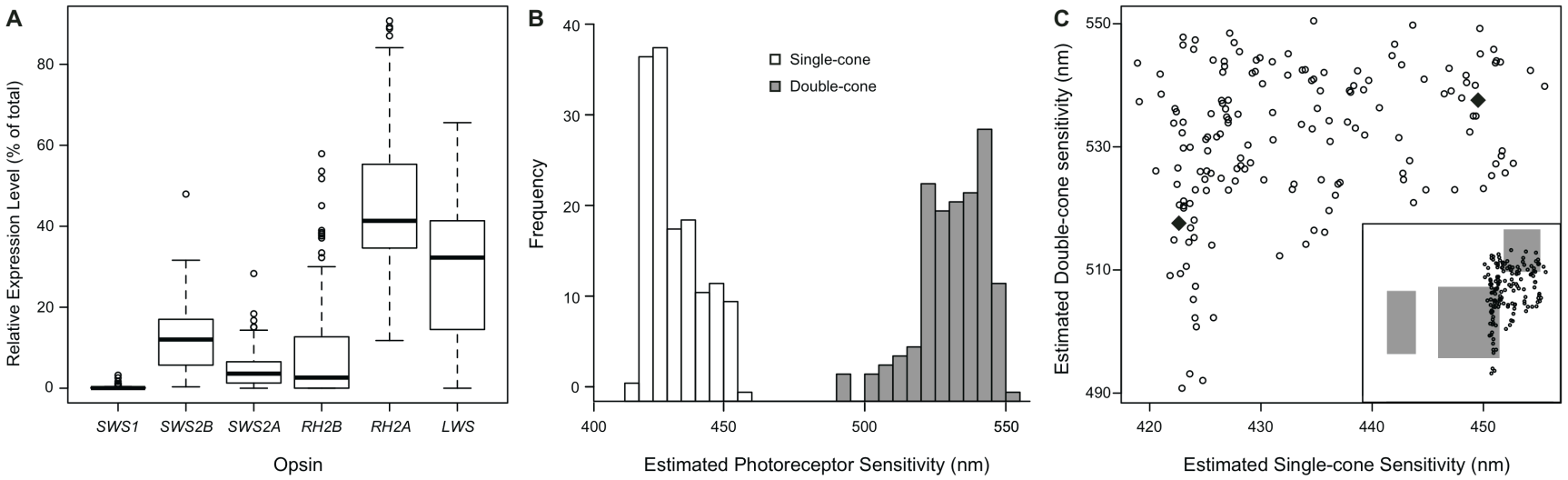
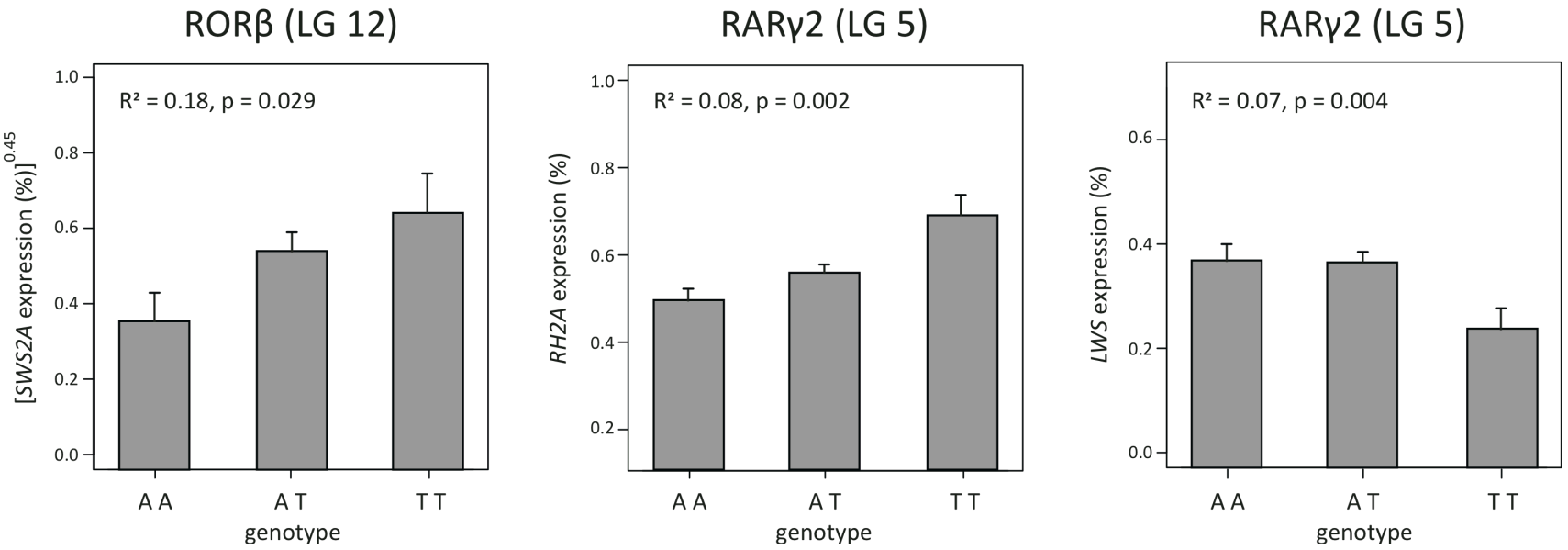


Figure 6-2. Expression QTL for opsin gene expression linked two *trans*-regulatory factors. Genotypic variation linked to RARγ-2 is associated with the expression of two opsins, *RH2A* and *LWS*. Genotypes refer to alleles derived from either *Aulonocara baenschi* (A) or *Tramitichromis intermedius* (T). Barplots show mean expression; error bars show standard error.



Chapter 7:

The Locus of Evolution from a Cichlids' Eye View:

Summary and Conclusion to Dissertation

One goal of evolutionary genetics is to elucidate the specific molecular genetic mechanisms that contribute to the evolution of phenotypes and the divergence of new species. These mechanisms include mutations within protein-coding genes as well as mutations within both *cis*- and *trans*-regulatory sequences (see Figure 1-1) [Chapter 1]. Although numerous studies have individually linked each of these molecular mechanisms to the evolution of specific traits (e.g., Rost et al. (2004), Hoekstra et al. (2006), Tishkoff et al. (2007), Gompel et al. (2004), Jones et al. (1988), Yvert et al. (2003)), none have examined how these factors collectively influence the evolution of a single trait. Such a comparison would reveal how these molecular genetic mechanisms may or may not differ in their contribution to phenotypic evolution. This has been the overarching goal of my dissertation. In this final chapter, I summarize the main conclusions from this dissertation and synthesize them into one final result, while also exploring the potential for future work. The results of this dissertation reveal that the molecular genetic basis of phenotypic evolution can be complex, involving mutations in protein-coding, *cis*-regulatory, and *trans*-regulatory DNA; and that mutations in these different types of DNA may make distinct contributions to phenotypic evolution. Future work will need to functionally validate the role of the putative regulatory polymorphisms identified here, and expand them to more detailed analyses within pairs of cichlid populations and species.

Overview of Dissertation Results

Chapters 2 – 4 (Hofmann et al. 2009; O'Quin et al. 2010; O'Quin et al. 2011) of this dissertation examined the diversity and evolutionary significance of opsin expression and sequence changes among nearly 100 cichlid species from Lakes Tanganyika, Malawi, and Victoria. In Chapters 2 and 3 (Hofmann et al. 2009; O'Quin et al. 2010), we found that evolutionary changes in opsin expression affected every single opsin gene among cichlids in Lakes Tanganyika and Malawi. These changes contributed to large (30 – 100 nm) predicted shifts in retinal sensitivity (Figures 2-2 and 3-2) [Chapters 2 and 3]. We also found that expression of the *SWS1* opsin was correlated with divergence in foraging preferences among cichlids in these two lakes (Figures 2-3 and 3-4) [Chapters 2 and 3]. In combination with evidence from cichlids and other teleost fish that *SWS1* expression can increase foraging success on zooplankton, these results suggests that at least some evolutionary changes in cichlid opsin expression are due to natural selection. Ancestral character state reconstruction revealed that the many similarities in opsin expression observed among cichlids in Lakes Tanganyika and Malawi have evolved independently (Figure 3-3) [Chapter 3], and in Chapter 4 (O'Quin et al. 2011) we show that this independent evolution likely involved repeated changes to cichlid development (Figure 4-1) [Chapter 4]. In contrast, evolutionary changes in opsin expression among cichlids from Lake Victoria primarily affected the *SWS2B* (violet-sensitive) opsin only (Figure 2-2) [Chapter 2]. However, these changes were also correlated with the ecology (ambient light intensity), once again suggesting that some opsin expression patterns

have evolved due to natural selection (Figure 2-3) [Chapter 2]. Finally, when we compared evolutionary changes in the regulation and protein-coding sequence of each opsin, we found that regulatory changes generally affect all opsins, while mutations within protein-coding sequences primarily affect opsins sensitive to the extremes of the visible light spectrum (*SWS1* [ultraviolet] and *LWS* [red]) (Figure 2-6; Table S3-4) [Chapter 2; Appendix 3]. Collectively, these results confirm the important role that regulatory divergence can play in the evolution of visual sensitivity among African cichlids—a point that was previously underappreciated (Carleton 2009; Hofmann and Carleton 2009).

In Chapters 5 and 6, we examined the genetic factors responsible for evolutionary changes in cichlid opsin expression. In Chapter 5, we performed a bioinformatic analysis of non-coding DNA surrounding the opsin genes of African cichlids in order to indentify putative *cis*-regulatory mutations that contribute to the evolution of opsin expression. We found that two cichlid species, *Metriaclima zebra* and *Oreochromis niloticus*, had diverged considerably in the identity and number of transcription factor and microRNA binding sites at two non-coding elements, five promoter sequences, and two 3' untranslated regions (Table 5-3) [Chapter 5]. We then resequenced several of these divergent sequences in a panel of 18 cichlids from Lake Malawi that differ in opsin gene expression. This association mapping panel revealed three single nucleotide polymorphisms upstream of the *SWS2A*, *RH2B*, and *LWS* opsins that were moderately correlated with divergence in cichlid opsin expression (Table 5-6) [Chapter 5]. Combined with the results of a previous quantitative genetic analysis of opsin expression in cichlids (Carleton et al. 2010),

these results reveal that *cis*-regulatory mutations likely contribute to variation in the expression of multiple cichlid opsins. In Chapter 6, we used an experimental cross of two Lake Malawi cichlids that differ in their patterns of opsin expression to determine what role mutations linked to fifteen transcription factors play in the *trans*-regulatory control of the opsins. We found that genetic variation linked to the transcription factors RORβ and RARγ-2 was significantly correlated with variation in the expression of the *SWS2A*, *RH2B*, and *LWS* opsins (Figure 6-2) [Chapter 6]. Combined once again with the results of a previous quantitative genetic analysis of opsin expression in these species (Carleton et al. 2010), these results reveal that mutations in at least two *trans*-regulatory factors also contribute to variation in cichlid opsin expression.

Synthesis

Taken together with previous studies, the results of my dissertation reveal that mutations in protein-coding, *cis*-, and *trans*-regulatory DNA all contribute to the evolution of color vision in cichlids. Evidence for each mechanism, as well as the opsins they affect, is summarized in Table 7-1. Below, I examine the identity and effect of specific protein-coding, *cis*-, and *trans*-regulatory mutations, and their overall impact on cichlid vision.

Protein-coding mutation

Protein-coding mutations predominately alter sensitivity of the *SWS1* and *LWS* opsins in cichlids. Individual mutations in the retinal binding pocket of these

absorbance (λ_{max}) of these opins (Yokoyama 2008), collectively shifting cichlid visual sensitivities by ~ 15 nm (Carleton et al. 2006; Jordan et al. 2006; Carleton 2009). One possible explanation for the relatively limited scope of protein-coding mutations in cichlids is that cichlids have no additional opsin paralogs sensitive to ultraviolet and red light. For this reason, regulatory mutations affecting these two opsins cannot alter sensitivity to these regions by replacing the *SWS1* or *LWS* opsins with opsins sensitive to shorter or longer wavelengths. Therefore, mutation to the *SWS1* and *LWS* opsins themselves is the only path left open to generate evolutionary change in ultraviolet and red sensitivity. Regulatory mutations that would increase the relative expression of the existing *SWS1* or *LWS* opsins could increase overall sensitivity to ultraviolet and red light, but would not make the organism sensitive to shorter or longer wavelength light (e.g., < 360 nm or > 561 nm). Such a change could only be achieved by mutations to the opsins themselves or by replacing the chromophore 11-*cis* retinal with 11-*cis*-3, 4 didehydroretinal. Among cichlids in Lake Victoria, several mutations within the *LWS* opsin protein have been demonstrated to increase sensitivity to longer wavelength red light (Terai et al. 2002; Terai et al. 2006; Seehausen et al. 2008) (see Table S2-6).

Cis-*regulatory mutation*

Mutations in *cis*-regulatory sequences contribute to variation in the expression of opsins sensitive to the middle portion of the visible light spectrum: *SWS2B* (violet), *SWS2A* (blue), *RH2B* (blue-green), and possibly *LWS* (red) as well (Table 7-1). *Cis*-regulatory mutations should affect the expression of only one or a few tightly linked opsins; therefore, the magnitude of such changes in cichlids is possibly limited to

incremental changes in the relative sensitivity to specific wavelengths of violet through green light (400 – 530 nm). In general, *cis*-regulatory mutations should only alter the expression or one or a few opsins. For this reason, mutations in *cis*-regulatory sequences are most likely responsible for the adaptive evolution of *SWS2B*/*SWS2A* (Hofmann et al. 2009) [Chapter 2] and *RH2A*/*LWS* (Carleton et al. 2005a) opsin expression found in cichlids from Lake Victoria, and possibly also the adaptive evolution of *SWS1* opsin expression found in cichlids from Lake Malawi and Tanganyika (Hofmann et al. 2009; O'Quin et al. 2010) [Chapters 2 and 3]. If mutations in *cis* are responsible for these changes, then the best candidates for the genetic control of these opsins are found within those divergent non-coding regions and putative *cis*-regulatory polymorphisms we identified in Chapter 5. For *SWS2B/SWS2A* expression, these candidates include the two divergent conserved non-coding elements (CNEs 6 and 7) and one insertion-deletion polymorphism (SWS2A-217) found upstream of the *SWS2A* opsin, as well as the divergent 3'-UTR region found downstream of the *SWS2B* opsin (Tables 5-3 and 5-6) [Chapter 5]. For *RH2A/LWS* expression, the polymorphism found upstream of the *LWS* locus control region (CNE10-570) in Chapter 5 is also an excellent candidate (Figure 5-4; Table 5-3). Finally, for *SWS1* expression, the divergent promoter region and conserved non-coding element (CNE 3) we identify surrounding the *SWS1* opsin are also good candidates (Table 5-3; Figure 5-3) [Chapter 5].

Trans-*regulatory mutation*

Finally, *trans*-regulatory mutations are associated with divergence in the expression of every opsin in cichlids except *RH2B* (Table 7-1). Specifically, one

trans-regulatory factor on cichlid linkage group (LG) 13 is associated with divergence in *SWS2B* and *SWS2A* expression in the same hybrid cross used in Chapter 6, and *SWS1* and *SWS2B* expression in another cross (Carleton et al. 2010). A second *trans*-regulatory factor on LG 4 is associated with divergence in *RH2A* and *LWS* opsin expression in this latter cross as well (Carleton et al. 2010). In Chapter 6, we identify a third *trans*-regulatory factor on cichlid LG 12 that is associated with divergence in *SWS2A* expression. A fourth potential *trans*-regulatory factor on LG 5 is associated with divergence in *RH2A* and *LWS* opsin expression. However, the position of this last *trans*-regulatory factor between two arrays of cichlid opsins means that we cannot yet determine whether its signal is part of a distinct *trans*-regulatory QTL or the existing *cis*-regulatory QTL previously identified on LG 5.

Since *trans*-regulatory factors are presumably diffusible transcription factors or non-coding RNAs that can regulate the expression of multiple genes, it is unsurprising that all putative *trans*-regulatory factors we identify are associated with the expression of multiple unlinked opsins. Within these genetic crosses, we also find that the individual *trans*-regulatory factors each explain a larger portion of the variance in opsin expression than the individual *cis*-regulatory factors or the potential *trans*-regulatory factors on LG 5. The *trans*-regulatory factors on LG 4, 12, and 13 each explain 16 – 33% of the variation in expression of the *SWS1*, *SWS2B*, *SWS2A*, *RH2A*, and *LWS* opsins (Carleton et al. 2010) [Chapter 6]. In contrast, the quantitative trait loci on LG 5 each explain 7 – 13% of the variation in the expression of the *SWS2B*, *SWS2A*, *RH2B*, *RH2A*, and *LWS* opsins. This difference in the variance of opsin expression explained by *trans*- (LG 4, 12, and 13) and putative *cis*-

(LG 5) regulatory factors is statistically significant (t-test for unequal variances: $t_{7.774} = 4.469$, $p = 0.002$) (Figure 7-1). This differences is perhaps biologically significant as well, since it suggest that *trans*-regulatory factors contribute more to the divergence in cichlid opsin expression than *cis*-regulatory factors, an observation that is contrary to the prevailing view that *cis*-regulatory mutations will explain much of the phenotypic diversity in animal form and other phenotypes.

Future Directions

Although this dissertation has shed light on the evolutionary and genetic basis of visual system and opsin expression divergence among African cichlids, much work remains to be done. Most importantly, perhaps, we should validate the candidate regulatory regions identified in Chapters 5 and 6 through additional fine-mapping and functional analyses in transgenic cichlids or cell lines. Pinpointing the specific mutations responsible for regulatory changes in cichlid opsin expression will allow us to address multiple unresolved questions regarding the evolutionary genetics of this trait.

First, for example, one could examine the causative loci for molecular signatures of natural selection. This analysis would reveal whether variation in the expression of opsins other than *SWS1*, *SWS2B*, and *LWS* may be due to natural selection. It is unclear whether changes in the expression of all opsins will be adaptive, since many opsins exhibit expression patterns that are developmentally (Carleton et al. 2008; O'Quin et al. 2011) and genetically (Carleton et al. 2010) correlated. These correlations could indicate that evolutionary changes in some

opsins are simply a passive result of their genetic and developmental correlation with *SWS1*, *SWS2B*, or *LWS* expression.

Second, one could also genotype the causative loci in additional genetic crosses or cichlid lineages. These analyses would reveal whether the repeated evolution of cichlid opsin expression is governed by the same loci in different cichlid species. Some recent analyses of parallel evolution have revealed that unique mutations often underlie similar phenotypic changes, even when these mutations affect the same gene (Prud'homme et al. 2006). However, it is unclear whether this will be the case for cichlid opsin expression. A recent genetic analysis of pigmentation evolution in Lake Malawi's cichlids revealed that a single hyplotype at the *Pax7* locus contributes to pigmentation differences in numerous cichlid genera, and that this haplotype arose only once (Roberts et al. 2009); therefore, it is unclear whether the same will be true for the loci that control opsin gene expression. One could also genotype each causative locus in pairs of cichlid populations, species, and genera. This analysis would reveal whether protein-coding, *cis*-regulatory, or *trans*-regulatory mutations contribute to the earliest stages of visual system divergence in cichlids. Carroll (2008) proposed that *cis*-regulatory mutations should contribute to the earliest stages of regulatory divergence; however, other authors have found that *cis*-regulatory mutations are often accompanied by mutations in *trans*-, possibly due to compensatory evolution (Wittkopp et al. 2004). An analysis of cichlid species pairs at different levels of evolutionary and phenotypic divergence could provide a definitive test of whether one mutation type or the other—or perhaps all three—govern the earliest stages of phenotypic evolution and speciation.

Third, in order to further dissect how all three molecular genetic mutations contribute to the evolution of a single phenotype, future work would do well to focus on the protein-coding and regulatory evolution of the *LWS* opsin. The results of this dissertation and other studies reveal that sensitivity to red light can be tuned by adaptive protein-coding (Terai et al. 2006) [Chapter 1] and regulatory changes in cichlids (Carleton et al. 2005a), including mutations both in *cis* and *trans* to *LWS* (Carleton et al. 2010) [Chapters 5 and 6] (Table 7-1). A focused study of this opsin would provide an unparalleled system in which to address the evolutionary genetics of mutation and adaptation. For example, to determine how *cis*- and *trans*-regulatory mutations individually contribute to variation in *LWS* expression and double-cone photoreceptor sensitivity, one could cross two cichlid populations or species that vary in *LWS* expression. It is possible to estimate the relative impact that these two regulatory mutations make to phenotypic divergence in *LWS* opsin expression and double-cone λ_{max} by then comparing the relative expression of the *LWS* opsin and the λ_{max} of long-wavelength-sensitive photoreceptors with the allele-specific expression of candidate regulatory alleles in the F_0 and F_1 (see Wittkopp et al. 2004). After dissecting the contribution that *cis*- and *trans*-regulatory mutations make to phenotypic divergence, heterologous expression essays can be used to measure what impact any *LWS* protein-coding mutations also segregating between these species have on double cone λ_{max}. These exciting functional analyses would allow us to quantify the contribution that different coding and regulatory mutations make to the evolution of a single phenotype.

Conclusion

If one goal of evolutionary genetics is to identify the "locus of evolution" (Hoekstra and Coyne 2007), then it is clear that this goal must now be expanded to include multiple "loci of evolution". Analysis of the evolution, development, and genetics of opsin expression in cichlids reveals that adaptive phenotypic evolution can be governed by multiple loci. And even for this relatively oligogenic trait, we find that these loci can include mutations of different types—protein-coding, *cis*-regulatory, and *trans*-regulatory. Future work in evolutionary genetics should aim to identify multiple loci underling complex traits, and elucidate the contributions that each makes to phenotypic divergence.

Tables

Table 7-1. Summary of molecular genetic factors that contribute to visual sensitivity evolution among African cichlid fishes.

Mutation	*SWS1*[1]	*SWS2B*	*SWS2A*	*RH2B*	*RH2A*	*LWS*
Protein-coding	✓ [2, 3, 7]	-	-	-	-	✓ [2, 8]
Cis-regulatory[2]	-	✓ [1]	✓ [1, 5]	✓ [1, 5]	-	✓ [5]
Trans-regulatory	✓ [1] (LG 13)	✓ [1] (LG 13)	✓ [1, 6] (LGs 12,13)	-	✓ [1, 6] (LGs 4, 5)	✓ [1, 6] (LGs 4, 5)
Adaptive Exp.	✓ [2, 3]	✓ [2]	-	-	-	✓ [4]

[1] Summary Key:

1: Carleton et al. (Carleton et al. 2010)

2: Chapter 2 (Hofmann et al. 2009)

3: Chapter 3 (O'Quin et al. 2010)

4: Carleton et al. (Carleton et al. 2005a)

5: Chapter 5

6: Chapter 6

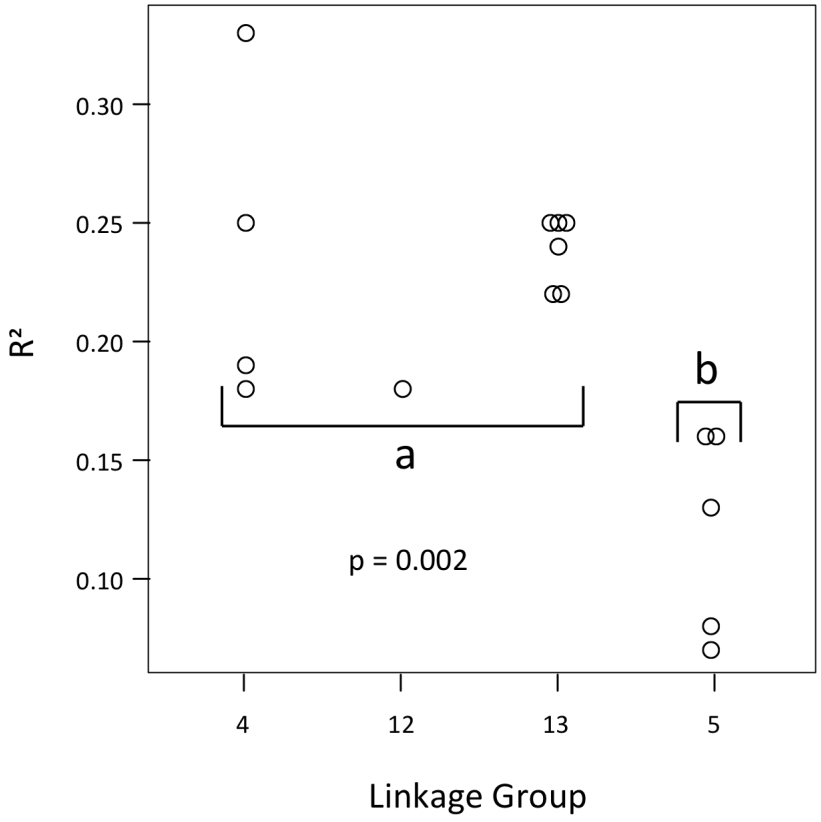
7: Smith and Carleton (2010)

8: Seehausen et al. (2008)

[2] All putative *cis*-regulatory mutations identified include QTL on LG5, as well as all polymorphisms upstream of the *SWS2A*, *RH2B*, and *LWS* opsins (also on LG5)

Figures

Figure 7-1. Putative *trans*-regulatory factors explain more variation in cichlid opsin expression than putative *cis*-regulatory factors. The coefficient of determination (R^2) for QTL on linkage groups 4, 5, and 13 are recorded from Carleton et al. (2010); others on linkage groups 12 and 5 are from Chapter 6.



Appendices

Appendix 1:

Supplemental Tables and Figures from Chapter 2

Supplementary Table S2-1. Sample size, relative opsin gene expression, estimated single- (SC) and double- (DC) cone sensitivity (λ_{max}), visual palette grouping ($k = 3$ clustering), foraging and habitat preference, and clade membership of all Lake Victoria and Lake Malawi cichlids used in Chapter 2.

Species	*n*	*SWS1*	*SWS2B*	*SWS2A*	*RH2B*	*RH2A*	*LWS*	SC λ_{max}	DC λ_{max}	$k = 3$ cluster	Forage[1]	Habitat[2]	Population / Clade
Paralabidochromis chilotes (1)	2	0.4	3.4	23.6	0.1	15.0	57.5	451	552	L	BI	CL	Ruti
Paralabidochromis chilotes (2)	1	0.5	2.5	23.5	0.1	12.2	61.2	451	554	L	BI	CL	Makobe
Pundamilia sp. "red head"	2	3.8	23.3	11.5	0.6	13.2	47.7	428	552	L	BI	TD	Zue
Neochromis omnicaeruleas	2	1.6	26.8	17.7	0.2	15.0	38.8	435	550	L	EA	CL	Ruti
Pundamilia azurea	2	1.0	2.0	46.2	0.1	14.4	36.3	453	550	L	EA	CL	Ruti
Lipochromis melanopterus	1	1.3	37.6	13.1	0.0	14.9	33.1	431	549	L	FH	CL	Makobe
Pundamilia nyerei (1)	5	1.0	2.2	25.7	0.1	16.8	54.0	450	552	L	ZP	VT	Python
Pundamilia nyerei (2)	2	2.9	18.4	14.0	0.2	9.0	55.4	432	555	L	ZP	CL	Makobe
Pundamilia nyerei (3)	1	0.7	20.3	13.6	0.1	8.7	56.5	436	555	L	ZP	CL	Senga
Pundamilia pundamilia (1)	4	1.4	22.8	14.3	0.1	12.4	49.0	434	553	L	BI	CL	Senga
Pundamilia pundamilia (2)	1	1.2	2.4	12.6	0.0	5.3	78.5	444	558	L	BI	VT	Kissenda
Cyathochromis obliquidens	4	15.3	0.6	0.2	28.7	13.9	41.4	363	524	S	PP	IN	mbuna
Cynotilapia afra	3	16.5	0.1	0.0	31.8	51.4	0.1	361	505	S	ZP	RK	mbuna
Genyochromis mento	2	15.9	1.6	0.2	22.4	48.6	0.5	367	504	S	FH	RK	mbuna
Labeotropheus fuelleborni	2	16.3	0.6	0.6	29.5	41.5	11.5	365	511	S	EA	RK	mbuna
Labeotropheus trewavasae	2	19.2	0.5	0.1	32.2	38.9	9.1	362	508	S	EA	RK	mbuna
Labidochromis sp. "blue bar"	1	3.9	7.7	2.3	20.0	42.3	23.7	412	522	M	PP	RK	mbuna
Labidochromis gigas	3	11.0	5.2	2.5	25.5	38.0	17.8	391	516	M	PP	RK	mbuna
Melanochromis auratus	3	0.6	21.8	0.1	33.7	39.4	4.4	423	504	M	EA	RK	mbuna

Continued

Supplementary Table S2-1 (continued).

Species	*n*	*SWS1*	*SWS2B*	*SWS2A*	*RH2B*	*RH2A*	*LWS*	SC λ_{max}	DC λ_{max}	*k* = 3 cluster	Forage[1]	Habitat[2]	Population / Clade
Melanochromis sp. "B&W johannii"	3	22.2	0.5	0.1	27.4	46.1	3.8	362	508	S	PP	IN	mbuna
Melanochrmos vermivous	3	2.1	18.4	0.1	33.9	42.8	2.7	418	504	M	PP	RK	mbuna
Melanochrmos parallelus	2	1.5	21.4	0.2	36.5	38.9	1.6	421	501	M	PP	IN	mbuna
Metriaclima aurora	5	17.3	1.0	0.2	32.3	45.5	3.6	365	505	S	EA	IN	mbuna
Metriaclima callainos	2	17.3	0.5	0.2	36.1	40.8	5.1	363	504	S	PP	RK	mbuna
Metriaclima livingstonii	1	15.6	0.0	0.0	33.9	46.4	4.1	360	505	S	PP	RK	mbuna
Petrotilapia nigra	5	21.9	0.4	0.6	19.5	45.1	12.4	364	517	S	PP	RK	mbuna
Metriaclima sp.	1	12.9	0.3	1.4	33.8	44.5	7.2	370	507	S	PP	RK	mbuna
Metriaclima zebra	3	13.4	0.9	0.4	30.3	52.4	2.6	367	507	S	ZP	RK	mbuna
Pseudotropheus microstoma	2	28.0	0.2	0.1	32.6	34.3	4.7	361	503	S	EA	IN	mbuna
Tropheops sp. "red cheek"	1	18.0	0.0	0.1	37.3	40.9	3.7	361	502	S	EA	RK	mbuna
Tropheops sp. "broad mouth"	1	22.8	0.1	0.0	32.0	43.4	1.7	360	504	S	EA	IN	mbuna
Tropheops gracillior	4	18.8	0.2	0.3	28.6	43.0	9.1	362	510	S	EA	IN	mbuna
Tropheops sp. "orange chest"	1	15.0	0.1	0.1	8.0	56.5	20.3	361	528	S	EA	RK	mbuna
Aristochromis christyi	2	4.9	17.6	0.8	25.4	47.4	3.5	412	509	M	FH	IN	utaka
Aulonocara hansbaenschi	2	1.4	8.5	0.5	19.0	55.2	15.3	418	520	M	BI	IN	utaka
Aulonocara sp.	1	2.0	18.1	0.9	35.7	40.8	2.5	420	502	M	BI	IN	utaka
Aulonocara sp. "blue fine"	4	6.4	18.3	1.3	22.9	29.8	21.3	411	519	M	BI	IN	utaka
Copadichromis eucinostomus	4	19.7	1.1	0.2	25.7	36.6	16.8	364	515	S	ZP	SD	utaka
Copadichromis jacksoni	1	15.0	0.2	0.0	6.7	50.3	27.7	361	532	S	ZP	PG	utaka
Cyrtocara moorii	1	3.4	0.1	6.9	0.0	36.9	52.8	424	546	L	BI	SD	utaka

Continued

Supplementary Table S2-1 (continued).

Species	*n*	*SWS1*	*SWS2B*	*SWS2A*	*RH2B*	*RH2A*	*LWS*	SC λ_{max}	DC λ_{max}	*k* = 3 cluster	Forage[1]	Habitat[2]	Population / Clade
Dimidiochromis compressiceps	1	1.2	1.3	8.5	0.0	45.7	43.3	442	542	L	FH	WD	utaka
Dimidiochromis kiwinge	1	16.6	1.4	2.3	11.1	58.5	10.1	375	522	S	FH	PG	utaka
Lethrinops aurita	4	10.8	3.5	0.3	23.3	22.4	39.7	377	527	S	ZP	SD	utaka
Maravichromis mola	2	2.7	12.2	0.4	16.5	49.4	18.8	414	522	M	BI	SD	utaka
Nimbochromis linni	2	9.3	16.1	0.8	33.4	35.2	5.1	403	503	M	FH	IN	utaka
Nimbochromis polysigma	2	19.0	4.1	0.5	33.1	31.5	11.8	373	507	S	FH	IN	utaka
Otopharynx heterodon	1	28.1	10.8	9.8	16.7	20.2	14.3	394	518	M	PP	IN	utaka
Ctenopharynx pictus	2	6.9	15.2	0.6	22.6	47.1	7.5	406	513	M	PP	IN	utaka
Placidochromis johnstoni	1	1.2	1.1	10.7	0.1	40.4	46.6	445	544	L	FH	WD	utaka
Placidochromis milomo	1	1.7	6.6	0.2	8.7	74.3	8.5	413	523	M	BI	RK	utaka
Protomelas annectens	1	11.0	27.7	0.2	30.5	29.5	1.0	407	499	M	BI	SD	utaka
Protomelas fenestraus	1	2.1	8.1	6.0	8.1	25.5	50.2	428	541	L	BI	IN	utaka
Protomelas similis	2	1.7	4.7	5.3	2.1	26.6	59.5	430	547	L	EA	WD	utaka
Protomelas spilonotus	1	16.0	16.7	7.3	12.0	32.6	15.4	405	523	M	ZP	RK	utaka
Protomelas taeniolatus	6	9.3	11.5	0.3	25.3	44.8	8.8	397	512	M	PP	RK	utaka
Stigmatochromis woodi	1	5.2	16.8	0.2	30.1	47.2	0.5	410	505	M	FH	PG	utaka
Taeniolatus praeorbitalis	1	9.6	6.0	3.9	8.8	52.0	19.7	399	528	M	BI	SD	utaka
Tramitichromis brevis	1	1.1	0.1	7.0	0.1	31.9	59.9	443	548	L	BI	SD	utaka
Trematocranus placodon	2	0.9	0.2	6.3	0.1	37.2	55.3	444	546	L	BI	IN	utaka
Tyrannochromis macrostoma	2	7.5	9.3	2.0	9.7	35.5	36.0	402	534	M	FH	RK	utaka
Tyrannochromis maculiceps	1	1.2	4.5	10.5	0.9	36.9	46.0	440	544	L	FH	RK	utaka

Continued

Supplementary Table S2-1 (continued).

Species	*n*	*SWS1*	*SWS2B*	*SWS2A*	*RH2B*	*RH2A*	*LWS*	SC λ_{max}	DC λ_{max}	*k* = 3 cluster	Forage[1]	Habitat[2]	Population / Clade
Hemitilapia oxyrhynchus	2	14.9	0.2	0.1	22.4	36.9	25.5	361	522	S	EA	WD	utaka
Pseudotropheus heteropictus	1	13.1	0.7	0.2	36.7	48.7	0.7	365	503	S	ZP	IN	mbuna
Rhamphochromis esox	1	1.6	5.3	12.1	3.6	20.7	56.7	417	515	M	FH	PG	utaka
Rhamphochromis sp.	1	1.4	2.9	1.5	15.5	78.7	0.0	439	547	L	FH	PG	utaka

[1] Foraging preference:
BI: Benthic invertebrates
EA: Epilithic algae
FH: Fish
PP: Phytoyplankton
ZP: Zooplankton

[2] Habitat preference (LV):
CL: Clear
TD: Turbid
VT: Very turbid

[2] Habitati preference (LM):
IN: Intermediate
PG: Pelagic
RK: Rock
SD: Sand
WD: Weed

Supplementary Table S2-2. Accession numbers for all opsins used in the opsin sequence diversity analysis from Chapter 2.

Species	*SWS1*	*SWS2B*	*SWS2A*	*RH2B*	*RH2Aα*	*RH2Aβ*	*LWS*	*RH1*
Pundamilia pundamilia (1)	AY673729	AY673709	AY673719	GQ422513	GQ422490	AY673699	AY673689	AY673739
Pundamilia pundamilia (2)	-	-	AY673760	-	-	AY673753	AY673748	AY673767
Pundamilia pundamilia (3)	-	-	AY673761	-	-	AY673754	AY90395	-
Pundamilia nyerei (1)	AY673728	AY673708	AY673718	GQ422522	GQ422482	AY673698	AY673688	AY673738
Pundamilia nyerei (2)	-	-	AY673762	-	-	AY673755	AY673749	AY673768
Pundamilia nyerei (3)	-	-	AY673763	-	-	AY673756	AY920394	-
Pundamilia azurea (1)	AY673730	AY673710	AY673720	GQ422504	GQ422487	AY673700	AY673690	AY673740
Pundamilia azurea (2)	-	-	AY673764	GQ422506	GQ422489	AY673757	AY673750	AY673769
Pundamilia sp. "red head" (1)	AY673732	AY673712	AY673722	GQ422505	GQ422485	AY673702	AY673691	AY673742
Pundamilia sp. "red head" (2)	-	-	AY673766	GQ422521	GQ422488	AY673759	AY673752	AY673771
Pundamilia luanso (1)	AY673731	AY673711	AY673721	GQ422510	GQ422484	AY673701	AY673692	AY673741
Pundamilia luanso (2)	-	-	AY673765	GQ422514	GQ422496	AY673758	AY673751	AY673770
Neochromis omnicaeruleus (1)	AY673735	AY673715	AY673725	GQ422502	GQ422481	AY673705	AY673695	AY673745
Neochromis omnicaeruleus (2)	-	-	-	GQ422503	GQ422492	-	-	-
Neochromis greenwoodi	AY673734	AY673714	AY673724	GQ422512	GQ422495	AY673704	AY673694	AY673744
Lipochromis melanopterus (1)	AY673733	AY673713	AY673723	GQ422507	GQ422486	AY673703	AY673693	AY673743
Lipochromis melanopterus (2)	-	-	-	GQ422509	GQ422491	-	-	-
Paralabidochromis chilotes	AY673736	AY673716	AY673726	GQ422511	GQ422493	AY673706	AY673696	AY673746
Paralabidochromis cyanus	AY673737	AY673717	AY673727	GQ422508	GQ422494	AY673707	AY673697	AY673747
Aulonocara hueseri	AY775100	AY775083	AY775074	GQ422516	GQ422480	AY775090	AY780517	AY775112
Aulonocara baenschi	GQ422525	GQ422528	GQ422527	GQ422520	GQ422499	GQ422500	GQ452104	GQ422474

Continued

Supplementary Table S2-2 (continued).

Species	*SWS1*	*SWS2B*	*SWS2A*	*RH2B*	*RH2Aα*	*RH2Aβ*	*LWS*	*RH1*
Cynotilapia afra	AY775104	AY775088	AY775079	GQ422517	GQ422497	AY775094	AY780521	AY775118
Labeotropheus fuelleborni	AF191223	AF247119	AF247115	-	-	AF247123	AF247127	AY775113
Labidochromis chismulae	AY775098	AY775064	AY775081	GQ422518	GQ422478	AY775069	AY780515	AY775120
Melanochromis auratus	AY775101	AY775084	AY775076	GQ422515	GQ422483	AY775091	AY780518	AY775115
Melanochromis vermivorus	DQ088643	DQ088640	DQ088637	DQ088646	DQ088631	DQ088634	DQ088628	GQ422472
Metriaclima zebra (1)	AF191219	AF317674	AF247114	DQ088652	DQ088651	AF247122	AF247126	AY775114
Metriaclima zebra (2)	AF191222	AF247118	-	-	-	-	-	-
Pseudotropheus acei	DQ088642	DQ088639	DQ088636	DQ088645	DQ088630	DQ088633	DQ088627	GQ422475
Copadichromis borleyi	AY775106	AY775065	AY775061	-	-	AY775071	AY780514	AY775121
Dimidiochromis compressiceps	AF191220	AF247117	AF247113	-	-	AF247121	AF247125	AY775059
Lethrinops parvidens	AY775102	AY775087	AY775077	-	GQ422477	AY775092	AY780519	AY775116
Mylochromis lateristriga	AY775105	AY775085	AY775075	GQ452103	GQ422476	AY775095	AY780522	AY775119
Stigmatochromis modestus	AY775107	AY775066	AY775080	-	-	AY775070	AY780523	AY775122
Tramitichromis intermedius (1)	DQ088644	DQ088641	DQ88638	DQ88647	DQ088632	DQ088635	GQ452105	GQ422473
Tramitichromis intermedius (2)	GQ422524	GQ422529	GQ422526	GQ422523	GQ422498	GQ422501	-	-
Tyrannochromis maculatus	AY775103	AY775086	AY775078	GQ422519	GQ422479	AY775093	AY780520	AY775117

Supplementary Table S2-3. Results of phylogenetic ANOVA comparing relative opsin expression and single and double cone sensitivity to foraging and habitat preference among cichlids from Lake Malawi using three phylogenetic hypotheses.

Tree	Opsin/Cone	Habitat		Foraging	
		$F_{4, 45}$	p-value	$F_{4, 45}$	p-value
Mitochondrial	*SWS1*	2.229	0.456	7.647	0.007**
	SWS2B	0.332	0.977	1.234	0.598
	SWS2A	1.468	0.656	1.957	0.367
	RH2B	4.294	0.066	3.323	0.141
	RH2A	3.127	0.214	0.755	0.810
	LWS	4.345	0.061	1.147	0.636
	Single Cone	2.179	0.493	9.065	0.002**
	Double Cone	4.084	0.085	2.067	0.363
Genera-level	*SWS1*	2.229	0.344	7.647	0.034
	SWS2B	0.332	0.951	1.234	0.761
	SWS2A	1.468	0.513	1.957	0.583
	RH2B	4.294	0.064	3.323	0.310
	RH2A	3.127	0.162	0.755	0.882
	LWS	4.345	0.056	1.147	0.789
	Single Cone	2.179	0.343	9.065	0.012**
	Double Cone	4.084	0.077	2.067	0.531
Star-level	*SWS1*	2.229	0.497	7.647	0.037
	SWS2B	0.332	0.976	1.234	0.760
	SWS2A	1.468	0.713	1.957	0.566
	RH2B	4.294	0.109	3.323	0.364
	RH2A	3.127	0.305	0.755	0.875
	LWS	4.345	0.099	1.147	0.763
	Single Cone	2.179	0.506	9.065	0.010**
	Double Cone	4.084	0.104	2.067	0.569

P-values following by ** are significant following Bonferroni correction for 8 hypothesis tests at an experiment-wise error rate of 10% ($\alpha = 0.10/8 = 0.0125$) (Quinn and Keough 2002).

Supplementary S2-4. Location, depth, Secchi disc readings, *SWS2B* quantum catch (QC), and *SWS2B* relative opsin expression for Lake Victoria taxa from Chapter 2.

Species	Location	Depth (m)	Secchi (cm)	*SWS2B* QC (%)	*SWS2B* exp (%)
Paralabidochromis chiloties (1)	Ruti	10	233	0.99	3.43
Paralabidochromis chilotes (2)	Makobe	2	225	4.29	2.48
Pundamilia sp. "read head"	Zue	2	150	4.29	23.29
Neochromis omnicaeruleas	Ruti	4	223	1.38	26.75
Pundamilia azurea	Ruti	12	223	0.05	2.04
Lipochromis melanopterus	Makobe	1	225	7.58	37.59
Pundamilia nyerei (1)	Python	4	98	0.04	2.25
Pundamilia nyerei (2)	Makobe	6	225	0.49	18.39
Pundamilia nyerei (3)	Senga	6	200	0.49	20.34
Pundamilia pundamilia (1)	Senga	2	200	4.29	22.79
Pundamilia pundamilia (2)	Kissenda	2	78	0.37	2.42

Supplementary Table S2-5. Summary of amino acid variation in opsin genes of cichlids from Lakes Malawi (LM) and Victoria (LV) presented in Chapter 2.

Segregating Sites	*SWS1*	*SWS2B*	*SWS2A*	*RH2B*	*RH2Aα*	*RH2Aβ*	*LWS*	*RH1*
Across all positions	26	16	15	16	22	32	24	22
Indels affecting codons[1]	-	-	+1	-	-	-	-1	-
Across all codons[2]	21	9	6*	10	10	22	15*	15
Non-synonymous (NS)	9	2	2	3	1	3	9	7
Across all TMR[3]	16	5	3	7	6	16	12	13
NS within TMR	7	1	1	1	1	3	8	7
Across all RBP[4]	5	3	1	2	1	3	5	3
NS within RBP	3	1	0	1	0	0	5	3
Unique to LV	3	1	1	1	1	0	6	9
Fixed between LM and LV	1	0	1	1	1	0	1	1

[1] + and – indicate addition or loss of whole codons due to insertion/deletions
[2] Amino Acids
[3] Transmembrane region
[5] Retinal binding pocket

Supplementary Table S2-6. Substitutions between amino acids with different physical properties that are located within the transmembrane regions of the cichlid opsins. Retinal binding pocket sites are shaded in gray. Amino acids sites that have previously been demonstrate to affect opsin spectral sensitivity are marked an asterisk (see text of Chapter 2). Parts A-B of this table cover the opsins *SWS1*, *SWS2B*, *SWS2A*, *RH2B*, *RH2Aα*, *RH2Aβ*, *LWS*, and *RH1*.

Part A. *SWS1*, *SWS2B*, *SWS2A*, *RH2B*, *RH2Aα*, and *RH2Aβ*.

	SWS1							*SWS2B*	*SWS2A*	*RH2B*	*RH2Aβ*			*RH2Aα*
Species	TM1	TM3	TM4		TM5		TM6	TM6	TM1	TM3	TM3	TM4	TM5	TM4
	37	114*	160	166	204	217	248	269*	39	124	107	151	218	151
Aulonocara hueseri	Y	S	T	G	T	S	K	A	A	S	A	T	V	T
Aulonocara baenschi	F	G	A	G	I	S	K	A	A	S	S	T	T/I	A
Cynotilapia afra	F	S	T	G	T	S	K	A	A	S	S	A	V	A
Labeotropheus fuelleborni	F	S	T	G	T	S	K	A	A	?	S	A	I	?
Labidochromis chismulae	F/Y	S	T	G	T	S	K	A	T	S	P	A	I	A
Melanochromis auratus	Y	S	T	G	T	F	K	A	A	S	S	A	T	A
Melanochromis vermivorus	F	S	T	G	T	S	K	A	A	S	S	A	V	A
Metriaclima zebra	F	S	T	G	T	S	K	A	T	S	S	A	I	A
Pseudotropheus acei	F	A	A	A	I	S	K	A	T	S	S	A	V/I	A
Copadichromis borleyi	Y	S	T	G	T	S	K	A	A	?	S	T	I	?
Dimidiochromis compressiceps	F	S	T	G	T	S	K	T	A	?	S	A	I	?
Lethrinops parvidens	F	S/T	T/A	G	T/I	S	K	A	A	?	P	A	I	A
Mylochromis lateristriga	F	S	T	G	T	S	K	A	A	A	S	T	I	A
Stigmatochromis modestus	F	S	T	G	T	S	R	A	A	?	A	T	T	?
Tramitichromis intermedius	F	S	T	G	T	S	K	T	A	S	P	A	I	A
Tyrannochromis maculates	F	S	T	G	T	S	K	A	A	A	S	T	I	A
Lipochromis melanopterus	F	A	A	G	I	S	K	T	A	S	P	A	I	A
Neochromis greenwoodi	F	A	A	G	I	S	K	?	A	S	P	A	I	A
Neochromis omnicaeruleus	F	A	A	S	I	S	K	T	A	S	P	A	I	A
Paralabidochromis chilotes	F	A	A	G	I	S	K	T	A	S	P	A	I	A
Paralabidochromis cyaneus	F	A	A	G	I	S	K	A	A	S	P	A	I	A
Pundamilia azurea	F	A	A	S	I	S	K	T	A	S	P	A	I	A
Pundamilia luanso	F	A	A	S	I	S	K	T	A	S	P	A	I	A
Pundamilia nyerei (1)	F	A	A	G	I	S	K	T	A	S	P	A	I	A
Pundamilia nyerei (2)	?	?	?	?	?	?	?	?	A	?	P	A	I	?
Pundamilia nyerei (3)	?	?	?	?	?	?	?	?	A	?	P	A	I	?
Pundamilia pundamilia	F	A	A	G	I	S	K	T	A	S	P	A	I	A
Pundamilia sp. "red head"	F	A	A	G	I	S	K	T	A	S	P	A	I	A

Supplementary Table S2-6 (continued).

Part B. *LWS* and *RH1*.

Species	*LWS*								*RH1*						
	TM1	TM3	TM4		TM5		TM6		TM1		TM4		TM5	TM6	
	40	123	155	164*	203	217	261*	262	42	158	163	166	213	298	299
Aulonocara hueseri	S/A	T/A	A	A	Y	A	Y	C	A	A	A	S	T	A	A
Aulonocara baenschi	S	T	A	A	Y	A	F/Y	C	A	A	A	S	T	A	A
Cynotilapia afra	A/A	T	A	A	Y	A	F	C	A	A	A	S	T	A	A
Labeotropheus fuelleborni	A	A	A	A	Y	A	Y	C	A	A	A	S	T	A	A
Labidochromis chismulae	A	A	A	A	Y	A	Y	C	A	A	A	S	T	A	A
Melanochromis auratus	A	A	A	S	Y	A	Y	C	A	A	A	S	T	A	A
Melanochromis vermivorus	A	A	A	S	Y	A	Y	C	A	A	A	A	T	S	S
Metriaclima zebra	A	A	A	A	Y	A	Y	C	A	A	A	S	T	A	A
Pseudotropheus acei	A	A	A	S	Y	A	Y	C	A	A	A	A	T	S	S
Copadichromis borleyi	A	A	A	A	Y	A	Y	C	A	A	A	A	I	S	S
Dimidiochromis compressiceps	A	A	A	A	Y	A	Y	C	A	A	A	A	I	S	S
Lethrinops parvidens	S	A	A	S	Y	A	Y	C	A	A	A	A	I	S	S
Mylochromis lateristriga	A	A	A	A	Y	A	Y	C	A	A	A	A	T	S	S
Stigmatochromis modestus	S	T	A	A	Y	A	Y	C	A	A	A	A	T	A	A
Tramitichromis intermedius	A	A	A	A	Y	A	Y	C	A	A	A	S	T	A	A
Tyrannochromis maculates	A	A	A	A	Y	A	Y	C	A	A	A	A	I	A	A
Lipochromis melanopterus	A	A	A	A	Y	A	Y	C	A	A/G	A	S	?	A	A
Neochromis greenwoodi	A	A	A	S	Y	A	Y	I	C	G	G	S	L	A	A
Neochromis omnicaeruleus	A	A	A	A	Y	A	Y	C	C	G	G	S	L	A	A
Paralabidochromis chilotes	A	A	A	A	F	T	Y	I	C	G	G	S	L	A	A
Paralabidochromis cyaneus	A	A	G	S	Y	A	Y	C	A	A	A	S	L	A	A
Pundamilia azurea	A	A	A	A	Y	T	Y	I	C	G	G	S	L	A	A
Pundamilia luanso	A	A	A	A	F	T	Y	I	C	G	G	S	L	A	A
Pundamilia nyerei (1)	A	A	A	A	Y	A	Y	C	C	G	G	S	L	A	A
Pundamilia nyerei (2)	A	A	A	A	F	T	Y	I	C	G	G	S	L	A	A
Pundamilia nyerei (3)	A	A	A	A	Y	A	Y	C	?	?	?	?	?	?	?
Pundamilia pundamilia	A	A	A	A	F	T	Y	I	C	G	G	S	L	A	A
Pundamilia sp. "red head"	A	A	A	A	Y	A	Y	C	A	A	A	S	L	A	A

Supplementary Table S2-7. Accession numbers for mtDNA sequences used to generate phylogenies for the comparative methods in Chapter 2.

Species	ND2	D-loop
Aristochromis christyi	EF585282	EF647535
Aulonocara baenschi	GQ422572	GQ422532
Aulonocara sp.	GQ422580	GQ422533
Aulonocara sp. "blue fin"	GQ422566	GQ422531
Copadichromis eucinostomus	EF585268	GQ422530
Copadichromis jacksoni	GQ422593	EF647580
Ctenopharynx pictus	GQ422587	GQ422547
Cyathochromis obliquidens	GQ422579	U90759
Cynotilapia afra	EF585264	AY911740
Cyrtocara moorii	AY930089	U01105
Dimidiochromis compressiceps	EF585267	EF647532
Dimidiochromis kiwinge	AF305322	AJ291408
Genyochromis mento	AF305297	U90779
Hemitilapia oxyrhynchus	EF585277	GQ422534
Labeotropheus fuelleborini	EF585259	U90774
Labeotropheus trewavasae	GQ422577	GQ422535
Labidochromis gigas	EF585276	GQ422538
Labidochromis sp. "blue bar"	GQ422573	GQ422537
Lethrinops aurita	GQ422586	GQ422539
Maravichromis mola	EF585274	GQ422540
Melanochromis auratus	AY930069	U01107
Melanochromis sp. "black-white johannii"	GQ422574	U01942
Melanochromis vermivorus	EF585270	GQ422541
Melanochromis parallelus	GQ422592	U01953
Metriaclima aurora	GQ422569	GQ422542
Metriaclima callainos	GQ422570	AF2136204
Metriaclima livingstonii	GQ422582	GQ422543
Metriaclima sp.	GQ422581	GQ422544
Metriaclima zebra	DQ093114	AY930025
Nimbochromis linni	EF585279	AY913941
Nimbochromis polystigma	EF585262	AJ291407
Otopharynx heterodon	EF585278	GQ422546
Petrotilapia nigra	GQ422567	GQ422548
Placidochromis johnstoni	EF585269	GQ422549

Continued

Supplementary Table S2-7 (continued).

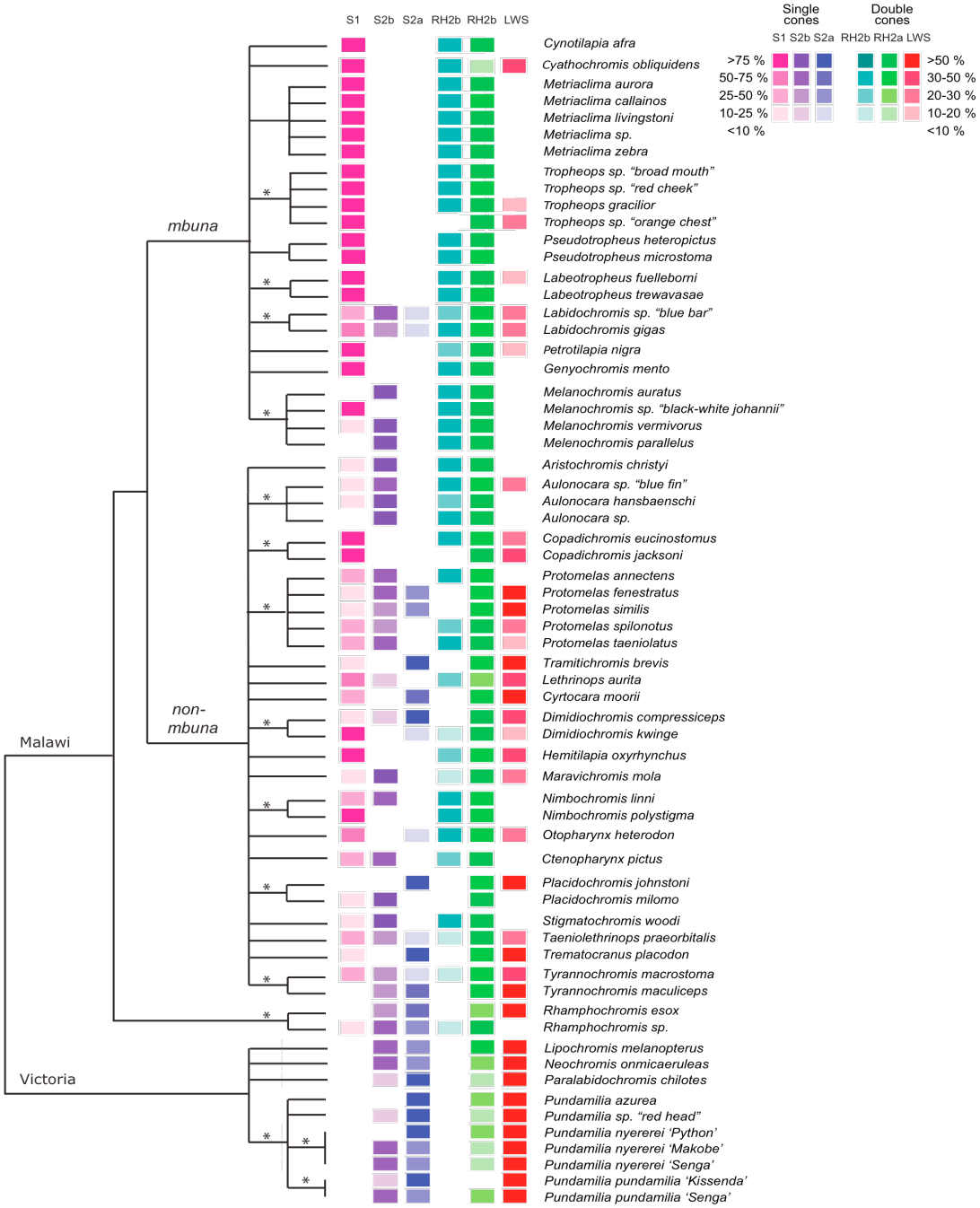
Species	ND2	D-loop
Placidochromis milomo	GQ422590	GQ422550
Protomelas annectens	GQ422575	AJ291414
Protomelas fenestratus	AF305301	GQ422551
Protomelas similis	GQ422585	GQ422552
Protomelas spilonotus	EF585253	GQ422553
Protomelas taeniolatus	AF305302	EF647546
Pseudotropheus heteropictus	GQ422584	GQ422554
Pseudotropheus microstoma	EF585258	GQ422555
Rhamphochromis esox	AF305252	AF298913
Rhamphochromis sp.	GQ422591	GQ422556
Stigmatochromis woodi	AF213626	GQ422557
Taeniolatus praeorbitalis	GQ422576	GQ422558
Tramitichromis brevis	AF305320	GQ422559
Trematocranus placodon	EF585261	GQ422560
Troheops gracillior	EF585260	GQ422562
Tropheops sp. "broad mouth"	GQ422589	GQ422561
Tropheops sp. "orange chest"	GQ422583	GQ422563
Tropheops sp. "red cheek"	GQ422568	GQ422564
Tyrannochromis macrostoma	EF585257	EF647537
Tyrannochromis maculicpes	GQ422571	GQ422565

Supplementary Table S2-8. Three distance-based validation statistics for the clusters of cichlid opsin expression and photoreceptor sensitivities presented in Chapter 2.

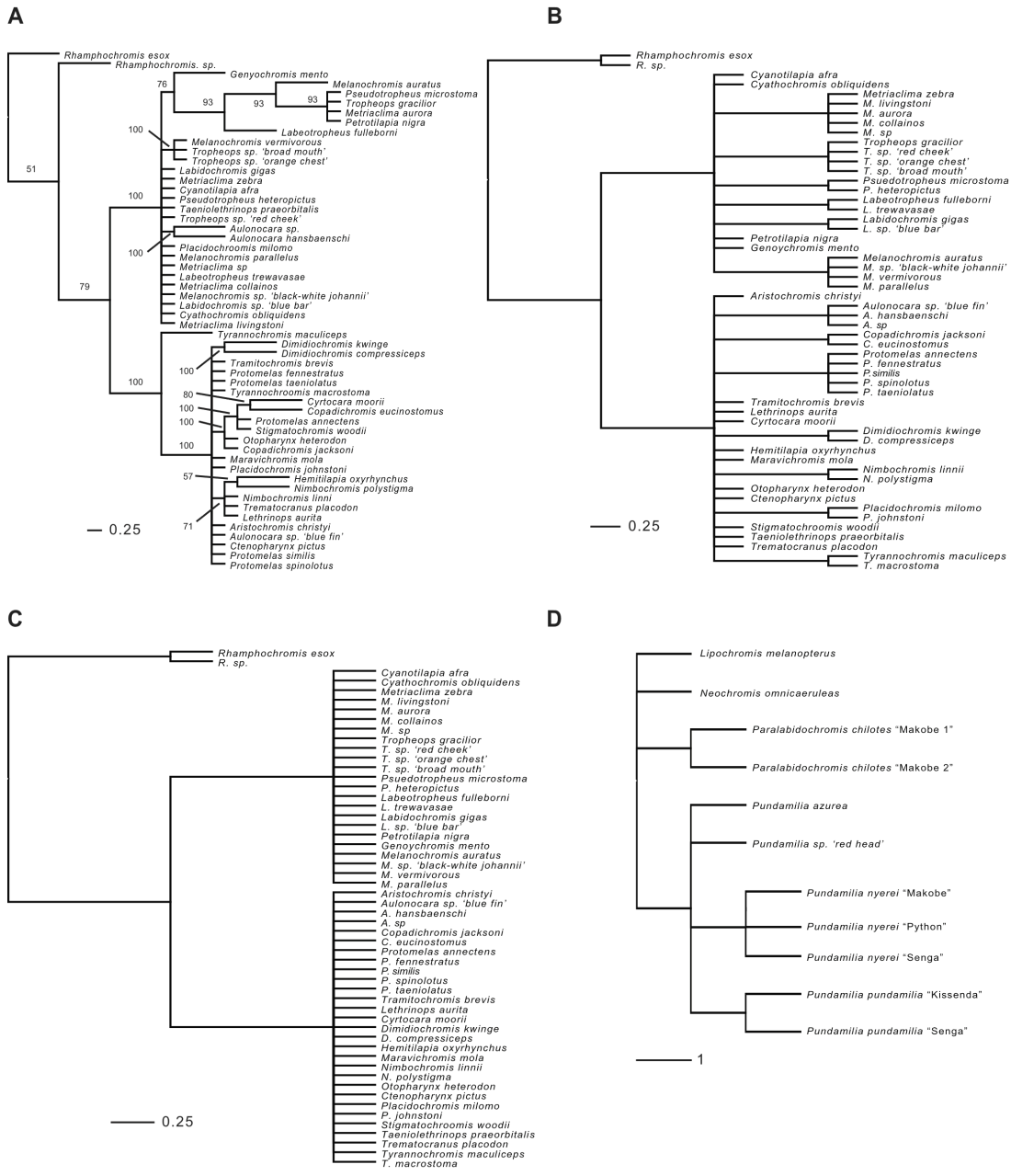
Variables	Index	Cluster size								
		2	3	4	5	6	7	8	9	10
Opsins	Connectivity	2.698**	3.270	7.174	13.794	26.208	33.822	42.737	44.785	51.397
	Dunn	0.196	0.311**	0.242	0.309	0.128	0.128	0.128	0.152	0.184
	Silhouette	0.460	0.576**	0.566	0.511	0.443	0.413	0.384	0.392	0.407
Photoreceptor Sensitivities	Connectivity	0.590**	0.590**	7.193	15.540	24.122	29.604	35.075	39.434	42.214
	Dunn	0.242	0.449**	0.153	0.199	0.116	0.140	0.140	0.140	0.184
	Silhouette	0.572	0.674**	0.620	0.558	0.486	0.496	0.509	0.515	0.542

** These values denote the optimal number of clusters indicated by each statistic.

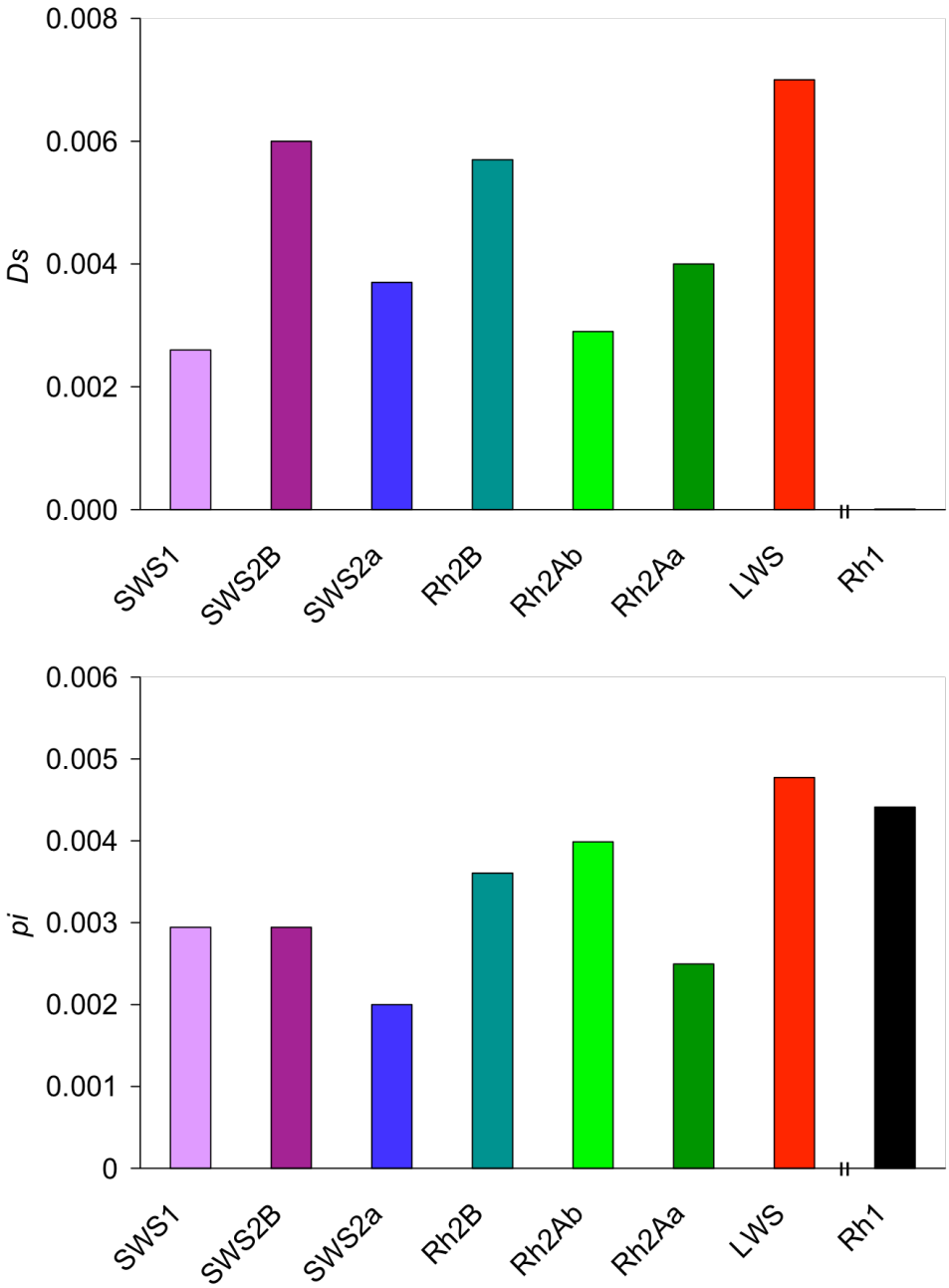
Supplementary Figure S2-1. Depiction of Lake Malawi and Lake Victoria cichlid opsin expression in a phylogenetic context.



Supplementary Figure S2-2. Phylogenies used for phylogenetically-control statistical methods (ANOVA and independent contrasts) in Chapter 2. (A - C) Phylogenetic trees for Lake Malawi cichlids. (A) Tree inferred from mitochondrial DNA, (B) tree based on taxonomic (genera-level) relationships, (C) tree based on rock-/sand-dwelling clade membership. (D) Phylogenetic tree for Lake Victoria cichlids based on taxonomic (genera-level) relationships.



Supplementary Figure S2-3. Synonymous substitution rates (D_S) and nucleotide diversity (*pi*) of each opsin gene for both Lake Malawi and Lake Victoria cichlids.



Appendix 2:

Supplementary Tables and Figures for Chapter 3

Supplementary Table S3-1. Primers used for the detection and sequencing of opsin and mtDNA in cichlids from Lake Tanganyika.

Opsin	Primer	Sequence
SWS1	F1a[1]	5'-GCGCGGAATTCAAAGAGCTCAGGGTCACAATG-3'
	R4[1]	5'-GCGCGCAAGCTTGCTCAGTCAACGCCCTCTTA-3'
	F2[1]	5'-GCGCGGAATTCGTGACCGCCTGGTCTTTG-3'
	R1a[1]	5'-GCGCGCAAGCTTCCCATGAACCAGGTGAAGG-3'
	R2[1]	5'-GCGCGCAAGCTTAGCAGCTGGGAGTAGCAGAA-3'
SWS2B	F3bb[1]	5'-GCGCGGAATTCTAGATTTTGATCGCAAACTCCAT -3'
	R2b[1]	5'-CCAAACAGAGGTGGAAGTGC-3'
	F2b[1]	5'-GCTTGTGGTCTCTTGCTGTGG-3'
	R1bb[1]	5'-GCGCGCAAGCTTCGGTTATTCACAACCCAGATG-3'
	F3c[1]	5'-TGCATGCAAGATTGAAGGAT-3'
	F1b[1]	5'-GATTATGGTGCTGGGCTTTC -3'
	R4b[1]	5'-CAGTATGCGAGCTGTCCAAA-3'
	R3[1]	5'-GCTTTCAGCATGAACAGCAG-3'
SWS2A	F1T[2]	5'-TACGGAGCTGCCAGAAGACT-3'
	R3T[2]	5'-GCAGAAGCAGAACAGGAACA-3'
	F2T[2]	5'-CCGCTCGGTAACTTTGTTTT-3'
	R5T[2]	5'-GCTCTCCTCCTCACCTCCTC-3'
RH2B	F1T[2]	5'-CTTGGGATGGAGGACTTGAG-3'
	R1T[2]	5'-AGCTTTTTATTAATTCAAGCTTTGG-3'
	F3T[2]	5'-AAAACCATCAGGATGCACA-3'
	R3T[2]	5'-TGTCTTTTATTTTAGGCGTTTCA-3'
	F4[1]	5'-TGCATCCCAACAGCAGGAC-3'
	R4[1]	5'-CAGGAAGGAGTATGGCTGGA-3'
RH2Aα	G2F1a[1]	5'-ACGCAGACTCAACTAAACAGC-3'
	R1a[1]	5'-GCGCGCAAGCTTGGACCATCCAAAGAGTGGAG-3'
	F2[1]	5'-GCGCGGAATTCGGTCACTTGTTGTCCTGGCT-3'
	G2R4[1]	5'-GGAAGCAATCATCAATGTCCA-3'
	R3[1]	5'-GCGCGCAAGCTTAGCACGTAGATAACAGGGTTGT-3'

Continued

Supplementary Table S3-1 (continued).

Opsin	Primer	Sequence
RH2Aβ[g]	G1F1a[1]	5'-GCGCGGAATTCGGGATATTCCATCAGCTGAAAC-3'
	G1R4[1]	5'-GCGCGCAAGCTTGCTTCTTAAATCCATTTGGCA-3'
	TDK-E[6]	5'-CCTGAAGTAGGAACCAGATG-3'
LWS	F0a[1]	5'-GCGCGGAATTCGGCTAACAGCTCAGGACCTC-3'
	R0[1]	5'-GCGCGCAAGCTTGCCCTCAAAGATACACATTGG-3'
	F0b[1]	5'-AAACTGTTTTCGACCAGTGT-3'
	R0a[1]	5'-CGAGACCGTTGGTGAAGACT-3'
	F1a[1]	5'-GCGCGGAATTCTTTGAGGGTCCCAATTACCA-3'
	R2[1]	5'-GCGCGCAAGCTTTCCACACAGCAAGGTAGCAC-3'
	F3[1]	5'-GCGCGGAATTCACTGGCCTCATGGACTGAAG-3'
	R5T[2]	5'-ATATTTATGCGGGAGCCACA-3'
ND2	Met[3]	5'-CATACCCCAACATGTTGGT-3'
	Trp[3]	5'-GAGATTTTCACTCCCGCTTA-3'
	ND2.2A[3]	5'-CTGACAAAAACTTGCCTT-3'
CYTB	L14725[4]	5'-CGAAGCTTGATATGAAAAACCATCGTTG-3 '
	H15573[5]	5'-AATAGGAAGTATCATTCGGGTTT-3 '
D-Loop	TDK-A[6]	5'-TTCCACCTCTAACTCCCAAAGCTAG-3'

[1] Parry et al. (2005)
[2] This study
[3] Kocher et al. (1995)
[4] Meyer et al. (1990)
[5] Taberlet et al. (1992)
[6] Lee et al. (1995)
[7] Sequencing of this opsin also used the primers R1a, F2, and R3 from *RH2Aα*.

Supplementary Table S3-2. Unique primer and probe combinations used to measure opsin gene expression in cichlids from Lake Tanganyika.

Combination	Primer/Probe	Opsins					
		SWS1	*SWS2b*	*SWS2a*	*RH2b*	*RH2a*	*LWS*
1	Forward	UV.Tang.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Cic.P	G3.Cic.P	G.Cic.P+G.Til.P	R.Cic.P
2	Forward	UV.Tang.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.Xeno.P	G3.Cic.P	G.Cic.P+G.Til.P	R.Cic.P
3	Forward	UV.Cic.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.P	G3.Cic.P	G.Til.P	R.Cic.P
4	Forward	UV.Cic.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.P	G3.Cic.P	G.Cic.P+G.Til.P	R.Cic.P
5	Forward	UV.Tang.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.P	G3.Cic.P	G.Til.P	R.Cic.P
6	Forward	UV.Tang.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.P	G3.Cic.P	G.Cic.P+G.Til.P	R.Cic.P
7	Forward	UV.Tang.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R+R.Tang.Nb.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.Nb.P	G3.Cic.P	G.Cic.P+G.Til.P	R.Cic.P

Continued

Supplementary Table S3-2 (continued).

Combination	Primer/Probe	Opsins					
		SWS1	*SWS2B*	*SWS2A*	*RH2B*	*RH2A*	*LWS*
8	Forward	UV.Tang.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.Nm.P	G3.Cic.P	G.Cic.P+G.Til.P	R.Cic.P
9	Forward	UV.Cic.F+UV.Tang.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.Pm.P	G3.Cic.P	G.Cic.P+G.Til.P	R.Cic.P
10	Forward	UV.Tang.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R+R.Tang.Xb.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.Xeno.P	G3.Cic.P	G.Cic.P+G.Til.P	R.Cic.P
11	Forward	UV.Cic.F+UV.Tang.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.P	G3.Cic.P	G.Til.P	R.Cic.P
12	Forward	UV.Tang.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.P	G3.Cic.P	G.Til.P+G.Tang.Gc.P	R.Cic.P
13	Forward	UV.Tang.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Tang.Nt.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.P	G3.Cic.P	G.Til.P	R.Cic.P
14	Forward	UV.Ov.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Tang.P	G3.Cic.P	G.Til.P+G.Tang.Ov.P	R.Cic.P

Continued

Supplementary Table S3-2 (continued).

Combination	Primer/Probe	Opsins					
		SWS1	*SWS2B*	*SWS2A*	*RH2B*	*RH2A*	*LWS*
15	Forward	UV.Cic.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Cic.P	G3.Cic.P	G.Cic.P+G.Til.P	R.Cic.P
Construct	Forward	UV.Cic.F	B2.Cic.F	B1.Cic.F	G3.Cic.F	G.Tang.F	R.Cic.F
	Reverse	UV.Cic.R	B2.Cic.R	B1.Tang.R	G3.Cic.R	G.Cic.R	R.Cic.R
	Probe	UV.Cic.P	B2.Cic.P	B1.Cic.P	G3.Cic.P	Cic.G.P	R.Cic.P

Supplementary Table S3-3. Relative opsin gene expression (% total ± s.e.), estimated single- (SC) and double- (DC) cone photoreceptor sensitivity (λ_{max}), and lens transmittance (T50) results for 28 cichlids species from Lake Tanganyika.

Species	*SWS1* (% total)	*SWS2b* (% total)	*SWS2a* (% total)	RH2b (% total)	*RH2a* (% total)	*LWS* (% total)	Est. SC λ_{max} (nm)	Est. DC λ_{max} (nm)	Lens T50 (nm)
Astatotilapia burtoni	0.69	9.27	12.2	0.26	31.3	46.2	439.23	545.40	-
Asprotilapia leptura	0.51±0.07	14.2±7.78	0.64±0.14	23.6±27.5	33.3±9.40	27.7±10.3	422.29	521.21	-
Benthochromis tricoti	0.67±1.03	1.53±1.82	1.12±1.52	41.8±8.74	54.7±11.7	0.19±0.13	421.45	501.03	397±0.35
Chalinochromis brichardi	0.80±0.12	6.31±5.02	0.28±0.02	55.7±2.82	22.2±0.15	14.7±2.22	417.39	498.33	-
Cyprichromis leptosoma	8.14±2.60	7.45±3.41	0.00±0.00	58.3±6.71	25.6±3.14	0.49±0.33	390.10	487.99	364±1.35
Eretmodus cyanostictus	10.4±2.11	1.96±0.75	0.05±0.04	35.6±4.04	17.8±3.73	34.3±2.14	370.32	517.17	349±7.66
Enantiopus melanogenys	0.21±0.15	28.8±4.51	0.49±0.41	36.9±5.93	13.6±3.97	20.0±10.2	423.09	507.10	-
Greenwoodichromis christyi	0.00±0.00	10.9±5.16	0.00±0.00	22.6±9.43	65.9±9.69	0.58±0.59	422.98	510.29	384±0.20
Julidochromis regani	0.97±0.14	2.48±1.31	0.08±0.08	67.5±3.88	25.9±5.34	3.10±2.82	406.38	488.56	-
Lobochilotes labiatus	0.00	3.39	1.70	0.20	22.2	72.5	433.99	551.93	409
Neolamprologus brichardi	0.98±0.39	9.89±5.60	0.01±0.02	63.3±6.25	25.2±3.57	0.63±0.59	417.38	487.03	-
Neolamprologus cunningtoni	0.43±0.17	21.8±3.44	0.02±0.02	42.3±6.72	34.7±3.86	0.72±0.52	421.82	495.59	385±9.33
Neolamprologus furcifer	0.58±0.20	11.3±3.73	0.56±0.42	35.8±27.3	29.7±11.1	22.1±15.2	421.55	511.75	-
Neolamprologus mondabu	0.43±0.20	20.2±9.54	3.05±2.05	22.3±16.0	18.5±6.85	35.5±16.7	426.09	525.79	-
Neolamprologus tretocephalus	0.13±0.10	7.16±1.22	72.4±10.6	8.38±0.85	11.5±10.6	0.46±0.38	452.87	502.82	-
Ophthalmotilapia ventralis	0.80±0.26	20.9±6.37	6.14±2.39	11.6±6.22	6.65±5.67	54.0±11.4	428.48	543.25	377±3.05
Petrochromis famula	0.31±0.13	13.6±1.54	0.01±0.01	45.5±2.56	27.9±3.38	12.7±4.86	421.60	501.67	375±2.40
Perissodus microlepis	0.74±0.22	14.5±0.55	0.00±0.00	56.5±5.77	20.6±12.3	7.62±7.26	419.95	492.41	-

Continued

Supplementary Table S3-3 (continued).

Species	*SWS1* (% total)	*SWS2b* (% total)	*SWS2a* (% total)	RH2b (% total)	*RH2a* (% total)	*LWS* (% total)	Est. SC λ_{max} (nm)	Est. DC λ_{max} (nm)	Lens T50 (nm)
Paracyprichromis nigrapinnis	18.6±2.53	0.10±0.58	1.30±0.53	1.43±2.13	25.0±9.17	53.5±5.21	366.55	547.53	373±17.5
Simochromis diagramma	0.01±0.01	2.79±0.83	1.52±0.32	0.37±0.28	22.2±3.11	73.1±3.69	434.52	551.84	400±7.06
Tanganicodus irsacae	8.96±0.65	1.09±0.78	0.11±0.15	36.2±4.13	27.4±2.94	26.2±6.42	367.80	513.56	356±7.47
Tropheus moori 'muzi'	0.00±0.00	3.52±0.19	1.69±0.56	0.19±0.08	23.0±0.98	71.6±1.38	433.66	551.60	405±2.78
Tropheus sp. mpimbwe	0.11±0.17	15.8±2.76	1.82±1.52	0.60±0.45	22.5±3.28	59.1±6.38	425.99	549.96	402±5.16
Xenotilapia bathyphila	0.25±0.06	27.4±3.83	0.13±0.10	37.7±9.71	7.81±2.14	26.8±3.90	422.60	510.49	-
Xenotilapia boulengeri	0.50±0.24	10.4±0.89	0.07±0.07	31.5±17.7	9.48±4.11	48.0±20.9	420.34	525.42	-
Xenotilapia flavipinnis	0.35±0.28	17.7±4.35	0.01±0.00	31.5±12.3	33.0±3.10	17.4±20.0	421.80	511.47	-
Xenotilapia ochrogenys	0.35±0.09	20.1±4.49	1.33±0.45	19.1±8.07	18.4±6.55	40.7±5.82	424.01	530.36	-
Xenotilapia spiloptera	0.49±0.15	3.56±2.97	0.05±0.06	42.3±59.7	52.9±57.4	0.65±0.76	415.90	500.75	-

Supplementary Table S3-4. Protein-coding sequence variation in ospisn genes of cichlids from Lake Tanganyika and *Oreochromis niloticus*. Parts A – C cover the opsin *SWS1*, *RH2Aβ*, and *LWS*.

Part A. *SWS1*

Species	TM 1				TM 2				TM 3			TM 4		TM 5				
	36[1]	52[2]	57	60	82	96	97	99	119	125	137	158	165	201	209	214	217	220
Astatotilapia leptura	V	T	I	V	S	V	S	M	V	S	M	A	I	T	V	I	T	I
Chalinochromis brichardi	A	T	I	V	S	L	A	M	V	A	L	A	I	T	V	I	S	I
Enantiopus melanogenys	V	T	I	L	S	V	S	M	T	A	L	G	I	T	V	F	A	I
Julidochromis regani	V	A	V	L	C	V	S	M	A	A	L	A	V	N	L	I	T	M
Neolamprologus brichardi	V	A	V	L	C	V	S	L	V	A	L	A	V	N	L	I	T	M
N. furcifur	V	A	V	L	C	V	S	M	A	A	L	A	V	N	L	I	T	M
N. tretocephalus	V	A	V	L	C	I	S	M	V	A	L	A	V	N	L	I	T	M
Perissodus microlepis	A	T	I	V	S	L	A	M	V	A	L	A	I	T	V	I	T	I
Xenotilapia bathyphila	A	T	I	L	S	V	S	M	A	A	L	A	I	T	V	I	T	I
X. spiloptera	V	T	I	L	S	V	S	M	A	A	L	A	I	T	V	I	A	I
Oreochromis niloticus	A	T	V	V	S	L	A	M	I	A	L	A	V	T	I	F	T	I

Species	TM6				TM7		
	252	261	270	276	290	293	298
Astatotilapia leptura	K	F	L	G	T	F	A
Chalinochromis brichardi	R	F	L	A	T	F	A
Enantiopus melanogenys	R	Y	L	G	S	L	S
Julidochromis regani	R	Y	L	A	S	L	A
Neolamprologus brichardi	R	F	L	A	T	F	A
N. furcifur	R	Y	L	A	S	L	A
N. tretocephalus	R	Y	L	A	S	L	A
Perissodus microlepis	R	F	V	A	T	F	A
Xenotilapia bathyphila	R	F	L	G	T	F	A
X. spiloptera	R	F	L	G	T	F	A
Oreochromis niloticus	R	F	L	A	I	F	S

Supplementary Table S3-4 (continued).

Part B. *RH2Aβ*

Species	TM 2	TM 3			TM 4				TM 5			TM 6			TM 7			
	99	107	109	112	151	158	162	165	203	214	218	255	259	263	286	290	304	309
Astatotilapia leptura	L	A	S	I	A	L	I	V	Y	F	I	A	M	I	L	L	I	M
Chalinochromis brichardi	V	A	S	V	A	L	I	V	Y	F	I	V	L	V	L	I	V	M
Enantiopus melanogenys	L	A	S	I	A	L	I	V	Y	F	I	V	M	V	L	I	I	M
Julidochromis regani	V	A	S	V	A	L	I	V	Y	F	I	V	L	V	L	I	V	M
Neolamprologus brichardi	V	A	S	V	A	L	I	V	Y	F	I	V	L	V	L	I	V	M
N. furcifur	L	A	F	I	A	F	V	M	Y	V	V	V	L	V	H	I	I	M
N. mondabu	V	A	S	V	A	L	I	V	Y	F	I	V	L	V	L	I	V	M
N. tretocephalus	L	A	S	V	A	L	I	V	Y	F	I	V	L	V	L	I	I	M
Perissodus microlepis	L	A	F	I	T	L	I	A	Y	C	V	V	M	I	L	I	V	M
Xenotilapia bathyphila	L	P	S	I	A	L	I	V	F	F	I	V	M	V	L	I	V	L
X. boulengeri	L	A	S	I	A	L	I	V	Y	F	I	V	M	I	L	I	I	M
X. flavipinnis	L	A	S	I	A	L	I	V	Y	F	I	V	M	I	L	I	I	?
X. ochrogenys	L	P	S	I	A	L	I	V	F	F	I	V	M	V	L	I	V	L
X. spiloptera	L	A	S	I	A	L	I	V	Y	I	I	V	M	V	L	I	I	M
Oreochromis niloticus	I	T	F	I	A	L	I	M	Y	Y	I	V	M	I	L	I	I	M

Continued

Supplementary Table S3-4 (continued).

Part C. *LWS*

Species	TM 1		TM 4				TM 5		TM 6	
	49	63	154	162	164[3]	172	209	218	259	270
Astatotilapia leptura	I	M	T	V	A	I	I	I	V	F
Chalinochromis brichardi	I	V	T	V	A	M	L	I	M	F
Enantiopus melanogenys	I	M	P	V	A	M	I	I	V	F
Neolamprologus furcifur	I	M	T	V	A	M	L	V	V	?
N. mondabu	I	M	T	V	S	M	L	V	V	V
N. tretocephalus	I	M	T	V	A	M	L	I	V	F
Perissodus microlepis	V	M	T	F	A	M	L	I	V	F
Xenotilapia boulengeri	I	M	T	L	?	I	I	I	V	F
X. bathyphila	I	M	T	V	S	I	I	V	V	F
X. flavipinnis	I	M	T	V	A	I	I	I	V	F
X. ochrogenys	I	M	T	V	S	I	I	V	V	F
Oreochromis niloticus	I	W	T	V	S	I	I	I	V	F

[1] Site numbers are the amino acid sites of bovine rhodopsin

[2] Sites highlighted in grey occur in or near a retinal-binding region and may therefore influence the spectral absorption of the opsin

[3] *LWS* site 164 (outlined in bold) is known to shift *LWS* sensitivity by 7 nm (Asenjo et al. 1994; Terai et al. 2006).

Supplementary Table S3-5. Results of Bayesian cluster analysis (k = 3 – 5 clusters) of opsin gene expression from 93 cichlids species from Lakes Tanganyika (LT), Malawi (LM), and Victoria (LV).

Species	Lake	k = 3	k = 4	k = 5	Visual Palette
Labidochromis gigas	LM	1	1	4	Middle[1]
Taeniolethrinops praeorbitalis	LM	1	1	4	Middle[1]
Otopharynx heterodon	LM	1	4	4	Middle[1]
Copadichromis eucinostomus	LM	1	1	1	Short[1]
Copadichromis jacksoni	LM	1	1	4	Short[1]
Cyanotilapia afra	LM	1	1	1	Short[1, 2]
Dimidiochromis kwinge	LM	1	1	4	Short[1]
Genyochromis mento	LM	1	1	1	Short[1]
Hemitilapia oxyrhynchus	LM	1	1	4	Short[1]
Labeotropheus fuelleborni	LM	1	1	1	Short[1, 7]
Labeotropheus trewavasae	LM	1	1	1	Short[1]
Melanochromis B&W johannii	LM	1	1	1	Short[1]
Metriaclima aurora	LM	1	1	1	Short[1]
Metriaclima callainos	LM	1	1	1	Short[1]
Metriaclima livingstoni	LM	1	1	1	Short[1, 2]
Metriaclima sp	LM	1	1	1	Short[1]
Metriaclima zebra	LM	1	1	1	Short[1, 3, 7]
Nimbochromis polystigma	LM	1	1	1	Short[1]
Petrotilapia nigra	LM	1	1	1	Short[1]
Pseudotropheus heteropictus	LM	1	1	1	Short[1]
Pseudotropheus microstoma	LM	1	1	1	Short[1]
Tropheops gracilior	LM	1	1	1	Short[1]
Tropheops sp. 'broadmouth'	LM	1	1	1	Short[1]
Tropheops sp. 'orangechest'	LM	1	1	4	Short[1]
Tropheops sp. 'redcheek'	LM	1	1	1	Short[1]
Cyathochromis oliquidens	LM	1	4	4	Short[1]
Lethrinops aurita	LM	1	4	4	Short[1]
Paracyprichormis nigrapinnis	LT	1	4	4	?[8]
Eretmodus cyanostictus	LT	1	4	4	Short[8]
Tanganicodus irascae	LT	1	4	4	Short[8]
Rhamphochromis sp.	LM	2	2	2	Long[1]
Aristochromis christyi	LM	2	2	2	Middle[1]
Aulonocara hansbaenschi	LM	2	2	2	Middle[1]
Aulonocara sp	LM	2	2	2	Middle[1]
Ctenopharynx pictus	LM	2	2	2	Middle[1]

Continued

Supplementary Table S3-5 (continued).

Species	Lake	$k = 3$	$k = 4$	$k = 5$	Visual Palette
Labidochroomis sp. 'bluebar'	LM	2	2	2	Middle[1]
Maravichromis mola	LM	2	2	2	Middle[1]
Melanochromis auratus	LM	2	2	2	Middle[1, 2]
Melanochromis parallelus	LM	2	2	2	Middle[1]
Melanochromis vermivorous	LM	2	2	2	Middle[1, 4]
Nimbochromis linni	LM	2	2	2	Middle[1]
Placidochromis milomo	LM	2	2	2	Middle[1]
Protomelas taeniolatus	LM	2	2	2	Middle[1, 2]
Stigmatochromis woodi	LM	2	2	2	Middle[1]
Aulonocara sp. 'bluefin'	LM	2	4	2	Middle[1]
Protomelas annectens	LM	2	4	2	Middle[1]
Protomelas spinolotus	LM	2	4	4	Middle[1]
Tyrannochromis macrostoma	LM	2	4	4	Middle[1, 2]
Benthochromis tricoti	LT	2	2	2	Middle[8]
Chalinochromis brichardi	LT	2	2	2	Middle[8]
Cyprichromis leptosoma	LT	2	2	2	Middle[8]
Greenwoodichromis christyi	LT	2	2	2	Middle[8]
Julidochromis regani	LT	2	2	2	Middle[8]
Neolamprologus brichardi	LT	2	2	2	Middle[8]
Neolamprologus cunningtoni	LT	2	2	2	Middle[8]
Neolamprologus furcifer	LT	2	2	2	Middle[8]
Perissodus microlepis	LT	2	2	2	Middle[8]
Petrochromis famula	LT	2	2	2	Middle[8]
Xenotilapia flavipinnis	LT	2	2	2	Middle[8]
Xenotilapia spiloptera	LT	2	2	2	Middle[8]
Asprotilapia leptura	LT	2	4	2	Middle[8]
Enantiopus melanogenys	LT	2	4	2	Middle[8]
Neolamprologus mondabu	LT	2	4	3	Middle[8]
Xenotilapia bathyphila	LT	2	4	2	Middle[8]
Xenotilapia boulengeri	LT	2	4	3	Middle[8]
Xenotilapia ochrogenys	LT	2	4	3	Middle[8]
Cyrotocara moorii	LM	3	3	5	Long[1]
Dimidiochromis compressiceps	LM	3	3	5	Long[1, 2, 7]
Placidochromis johnstoni	LM	3	3	5	Long[1]
Protomelas fenestratus	LM	3	3	5	Long[1]
Protomelas similis	LM	3	3	5	Long[1]

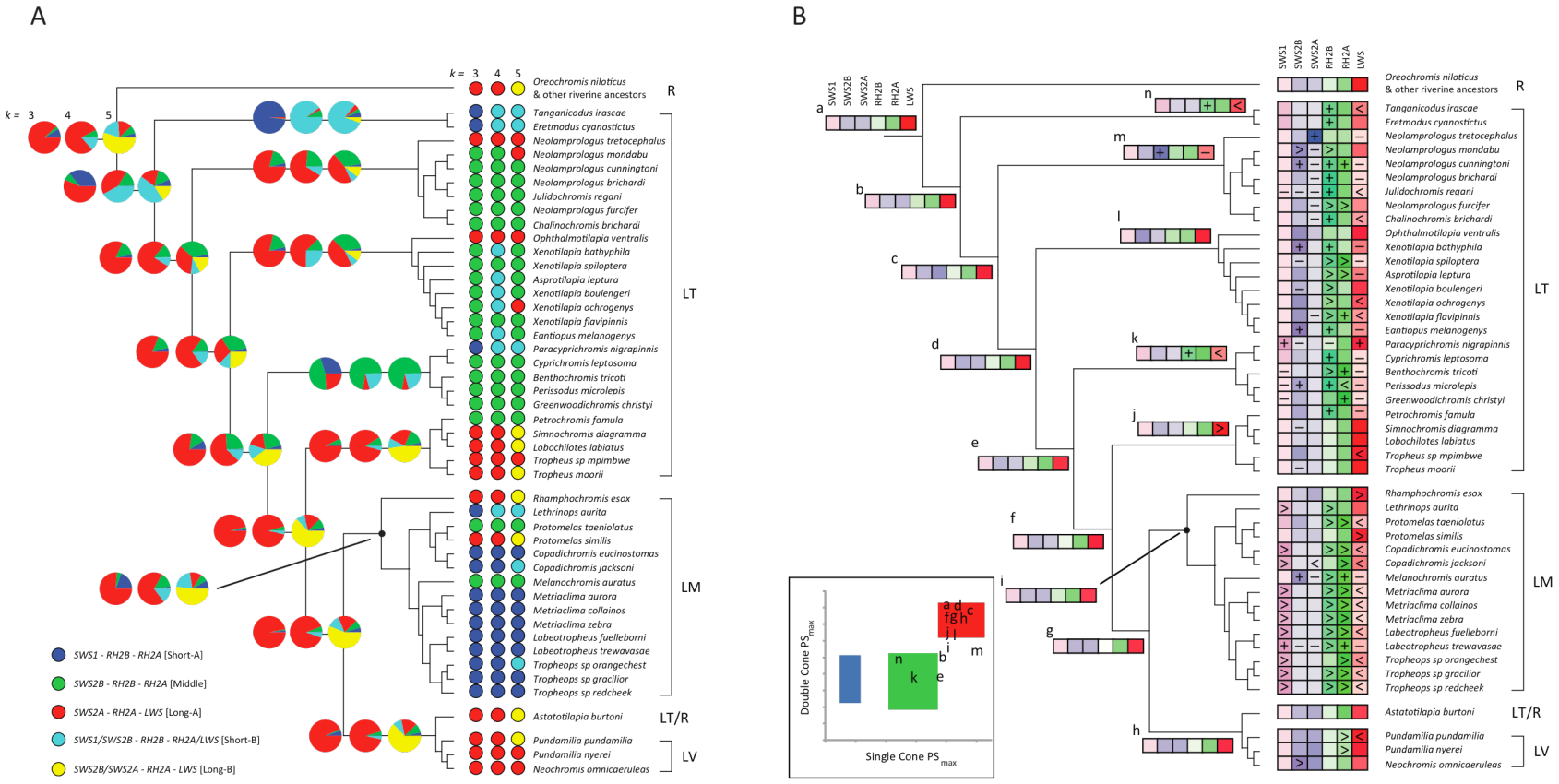
Continued

Supplementary Table S3-5 (continued).

Species	Lake	$k = 3$	$k = 4$	$k = 5$	Visual Palette
Tramitochromis brevis	LM	3	3	5	Long[1]
Trematocranus placodon	LM	3	3	5	Long[1]
Tyrannochromis maculiceps	LM	3	3	5	Long[1]
Rhamphochromis esox	LM	3	3	5	Middle[1]
Neolamprologus tretocephalus	LT	3	3	3	?[8]
Astatotilapia burtoni	LT	3	3	5	Long[5, 8]
Lobochilotes labiatus	LT	3	3	5	Long[8]
Ophthalmoliapia ventralis	LT	3	3	3	Long[8]
Simochromis diagramma	LT	3	3	5	Long[8]
Tropehus moorii	LT	3	3	5	Long[8]
Tropheus sp. mpimbwe	LT	3	3	3	Long[8]
Lipochromis melanopterus	LV	3	3	3	Long[1]
Neochromis omnicaeruleas	LV	3	3	3	Long[1]
Paralabidochromis chilotes	LV	3	3	3	Long[1]
Pundamilia azurea	LV	3	3	3	Long[1]
Pundamilia nyerei 'Makobe'	LV	3	3	3	Long[1, 6]
Pundamilia nyerei 'Python'	LV	3	3	3	Long[1, 6]
Pundamilia nyerei 'Senga'	LV	3	3	3	Long[1, 6]
Pundamilia pundamilia 'Kissenda'	LV	3	3	5	Long[1]
Pundamilia pundamilia 'Senga'	LV	3	3	3	Long[1]
Pundamilia sp. 'redhead'	LV	3	3	3	Long[1]
Oreochromis niloticus	R	3	3	5	Long[7]

[1] Hofmann et al. (2009)
[2] Jordan et al. (2006)
[3] Carleton et al. (2000)
[4] Parry et al. (2005)
[5] Fernald and Liebman (1980)
[6] Carleton et al. (2005)
[7] Carleton et al. (2008)
[8] This Study

Supplementary Figure S3-1. Parallel evolution of opsin gene expression in 47 African cichlids fishes from Lakes Tanganyika (LT), Malawi (LM) and Victoria (LV), as well as the rivers (R), continued. (A) Pie charts illustrate the results of Bayesian ancestral state reconstruction and show the relative posterior probability that the ancestor expressed each of three opsin expression palettes determined by clustering taxa into k = 3, 4, and 5 clusters. In each case, a long wavelength-like opsin palette (red or yellow) is inferred to be the ancestral state for most African cichlid lineages, including the Haplochromini (LM and LV). States at the tips indicate several parallel shifts to the short (blue and light-blue) and middle (green) wavelength palettes among cichlids in LT and LM from ancestors that each expressed the long wavelength palette. (B) Heat maps illustrate the results of Maximum Likelihood continuous character state reconstruction of each opsin's expression profile individually across the cichlid phylogeny. Color indicates which opsin is being reconstructed while intensity indicates the amount of expression (white indicates little or no expression and a deep color indicates high expression) for that opsin. Pluses and minuses indicate expression values that are significantly greater or smaller than the previous ancestor, except for at the tips of the phylogeny, where each taxa is compared to the nearest reconstructed ancestral state at the base of each clade. Inset) Single and double cone PS_{max} estimated from the reconstructed opsin expression profiles of each ancestral node lettered a - n. Like the Bayesian ancestral state reconstructions, the results of this analysis indicate that the ancestors of most cichlid lineages express opsins consistent with the long wavelength palette (*SWS2A-RH2A-LWS* , red inset box), followed by numerous parallel shifts in opsin expression among cichlids in each lake.


A
B
k = 3 4 5
Oreochromis niloticus & other riverine ancestors
R
Tanganicodus irascae
Eretmodus cyanostictus
Neolamprologus tretocephalus
Neolamprologus mondabu
Neolamprologus cunningtoni
Neolamprologus brichardi
Julidochromis regani
Neolamprologus furcifer
Chalinochromis brichardi
Ophthalmotilapia ventralis
Xenotilapia bathyphila
Xenotilapia spiloptera
Asprotilapia leptura
Xenotilapia boulengeri
Xenotilapia ochrogenys
Xenotilapia flavipinnis
Eantiopus melanogenys
Paracyprichromis nigrapinnis
Cyprichromis leptosoma
Benthochromis tricoti
Perissodus microlepis
Greenwoodichromis christyi
Petrochromis famula
Simnochromis diagramma
Lobochilotes labiatus
Tropheus sp mpimbwe
Tropheus moorii
LT
Rhamphochromis esox
Lethrinops aurita
Protomelas taeniolatus
Protomelas similis
Copadichromis eucinostomas
Copadichromis jacksoni
Melanochromis auratus
Metriaclima aurora
Metriaclima collainos
Metriaclima zebra
Labeotropheus fuelleborni
Labeotropheus trewavasae
Tropheops sp orangechest
Tropheops sp gracilior
Tropheops sp redcheek
LM
Astatotilapia burtoni
LT/R
Pundamilia pundamilia
Pundamilia nyerei
Neochromis omnicaeruleas
LV
SWS1 - RH2B - RH2A [Short-A]
SWS2B - RH2B - RH2A [Middle]
SWS2A - RH2A - LWS [Long-A]
SWS1/SWS2B - RH2B - RH2A/LWS [Short-B]
SWS2B/SWS2A - RH2A - LWS [Long-B]
SWS1 SWS2B SWS2A RH2B RH2A LWS
Double Cone PS max
Single Cone PS max

Appendix 3:

Supplementary Tables for Chapter 4

Supplementary Table S4-1. Relative opsin gene expression (% of total opsin expression) and predicted photoreceptor sensitivity (PS_{max}, nm) of single- (SC) and double- (DC) cones measured for all samples reported in Chapter 4, including *Oreochromis niloticus*.

Species	dpf	*SWS1*	*SWS2B*	*SWS2A*	*RH2B*	*RH2A*	*LWS*	SC PS_{max}	DC PS_{max}
Astatotilapia	14	7.39	11.56	0.00	0.01	25.01	56.03	398.45	549.26
burtoni	14	22.24	7.74	0.00	0.01	22.41	47.60	376.26	548.82
	14	19.12	8.29	0.00	0.01	21.00	51.58	379.05	549.99
	28	2.68	19.12	1.87	0.01	27.57	48.75	418.49	547.26
	28	14.42	12.92	0.17	0.01	25.43	47.04	390.20	547.65
	28	12.14	12.75	0.14	0.03	26.65	48.29	392.64	547.46
	42	1.19	6.49	4.14	0.00	37.19	50.99	428.24	544.97
	42	1.23	7.51	5.50	0.01	35.37	50.39	430.31	545.32
	42	2.11	3.88	1.67	0.01	41.81	50.53	412.84	543.79
	42	4.91	10.36	0.75	0.01	34.18	49.79	405.25	545.53
	42	0.96	6.23	5.27	0.00	41.09	46.44	432.09	543.16
	42	2.74	9.56	1.80	0.00	35.99	49.91	414.98	545.08
	56	1.08	2.40	6.04	0.00	41.34	49.14	436.80	543.63
	56	1.34	3.35	2.34	0.00	38.85	54.11	421.94	545.12
	56	0.10	0.60	0.00	4.42	15.38	79.50	414.19	551.15
	56	1.38	5.04	3.52	0.01	41.52	48.53	425.94	543.48
	56	1.24	5.42	43.96	0.00	19.97	29.41	450.12	545.63
	56	2.96	11.67	0.00	0.00	33.02	52.35	410.25	546.30
	70	0.96	1.19	4.71	0.01	39.23	53.90	436.82	544.99
	70	1.01	2.01	3.10	0.01	36.76	57.11	429.28	546.11
	70	1.10	1.69	2.01	0.01	41.11	54.08	422.35	544.58
	70	0.02	2.83	0.00	0.00	40.67	56.47	422.65	545.09
	70	0.02	1.43	0.00	0.01	44.59	53.95	422.14	543.80
	70	0.17	3.68	0.00	0.00	39.11	57.03	420.18	545.54
	Adt[1]	0.69	9.27	12.20	0.26	31.30	46.20	439.21	545.41
Melanochromis	10	15.87	22.04	0.17	14.36	36.19	11.37	396.90	518.15
auratus	10	22.59	20.79	0.29	17.91	27.47	10.96	390.63	514.18

Continued

Supplementary Table S4-1 (continued).

Species	dpf	*SWS1*	*SWS2B*	*SWS2A*	*RH2B*	*RH2A*	*LWS*	SC PS_{max}	DC PS_{max}
Melanochromis	15	10.15	25.24	0.05	22.11	33.16	9.29	405.00	511.00
auratus	15	10.54	21.64	0.06	22.16	37.23	8.36	402.48	511.00
	21	2.69	28.76	0.18	21.42	32.81	14.14	417.83	514.89
	21	3.34	28.67	0.22	16.96	35.42	15.39	416.70	518.86
	26	1.99	31.16	0.31	23.65	15.23	27.65	419.55	520.66
	26	2.47	26.75	0.30	21.86	25.78	22.84	418.06	519.50
	31	2.01	31.00	0.44	18.64	36.72	11.19	419.65	515.10
	31	2.27	34.00	0.47	21.20	31.20	10.86	419.52	512.44
	37	14.82	18.99	6.23	9.70	33.47	16.80	404.82	525.40
	37	8.34	21.40	5.46	18.41	30.16	16.23	413.18	518.03
	44	1.15	25.15	0.32	21.17	39.92	12.28	420.68	514.65
	44	1.72	24.89	0.29	24.03	33.35	15.72	419.34	514.41
	50	1.98	23.13	0.07	21.23	41.11	12.49	418.13	514.87
	50	1.67	22.28	0.19	19.75	41.93	14.19	418.91	516.83
	58	2.00	23.58	0.33	20.04	39.26	14.79	418.57	516.79
	58	2.58	24.74	0.33	20.32	36.83	15.19	417.53	516.65
	65	2.84	27.09	0.06	23.29	32.56	14.16	417.10	513.72
	65	2.71	26.63	0.21	24.87	33.49	12.09	417.45	511.52
	72	0.55	23.37	0.65	20.94	41.69	12.80	422.45	515.29
	72	1.66	22.83	0.55	23.89	38.63	12.43	419.54	513.05
	Adt[2]	0.60	21.80	0.10	33.70	39.40	4.40	421.47	502.98
Oreochromis	7	0.35	0.01	0.00	0.24	0.10	0.30	362.42	521.07
niloticus[3]	9	0.30	0.02	0.00	0.27	0.16	0.25	364.27	516.15
	12	0.29	0.03	0.00	0.30	0.15	0.23	365.50	512.54
	14	0.36	0.02	0.00	0.23	0.21	0.17	364.11	514.17
	16	0.20	0.05	0.00	0.19	0.17	0.39	373.61	529.01
	18	0.29	0.05	0.00	0.30	0.17	0.19	369.65	510.96
	20	0.27	0.04	0.00	0.28	0.25	0.16	368.92	511.24
	22	0.23	0.04	0.00	0.30	0.24	0.21	368.83	512.74
	24	0.15	0.09	0.00	0.22	0.27	0.28	384.27	522.06
	26	0.27	0.02	0.00	0.25	0.18	0.28	365.24	519.96
	28	0.20	0.04	0.00	0.30	0.23	0.23	371.06	513.82
	31	0.18	0.07	0.00	0.21	0.23	0.31	378.71	524.08
	33	0.11	0.09	0.00	0.20	0.30	0.30	389.83	524.34
	36	0.03	0.18	0.02	0.05	0.20	0.52	418.36	544.76

Continued

Supplementary Table 4-1 (continued).

Species	dpf	*SWS1*	*SWS2B*	*SWS2A*	*RH2B*	*RH2A*	*LWS*	SC PS_{max}	DC PS_{max}
Oreochromis	43	0.06	0.09	0.01	0.14	0.17	0.52	403.59	537.31
niloticus[3]	45	0.06	0.13	0.00	0.17	0.22	0.43	405.72	532.10
	62	0.01	0.12	0.05	0.00	0.12	0.69	429.95	553.92
	64	0.01	0.12	0.07	0.00	0.11	0.69	433.58	554.75
	114	0.00	0.07	0.07	0.00	0.16	0.69	438.14	552.76
	116	0.00	0.05	0.09	0.00	0.20	0.66	441.99	551.26
	156	0.00	0.04	0.11	0.00	0.12	0.72	447.81	554.53
	196	0.00	0.04	0.10	0.00	0.11	0.75	447.52	555.35
	224	0.01	0.01	0.06	0.00	0.05	0.86	443.70	557.79
	Adt	0.00	0.00	0.08	0.00	0.09	0.83	453.59	556.39

[1] *A. burtoni* adult opsin expression from O'Quin et al. (2010)
[2] *M. auratus* adult opsin expression from Hofmann et al. (2009)
[3] *O. niloticus* developmental and adult opsin expression from Carleton et al. (2008)

Appendix 4:

Supplemental Tables and Figures from Chapter 5

Supplementary Table S5-1. GenBank accession numbers for all Lake Malawi cichlid sequences generated in Chapter 5.

Species	*LWS*	*RH2B*	*SWS2A*	*SWS2B*	*SWS1*	CNE-10	*LWS*-UTR	*SWS2B*-UTR
Rhamphochromis esox	JF262688	JF262706	JF262742	JF262760	JF262724	JF262670	HQ993484	HQ993470
Trematocranus placodon	JF262689	JF262707	JF262743	JF262761	JF262725	JF262671	HQ993485	HQ993471
Placidochromis johnstoni	JF262690	JF262708	JF262744	JF262762	JF262726	JF262672	HQ993482	HQ993468
Dimidiochromis compressiceps	JF262691	JF262709	JF262745	JF262763	JF262727	JF262673	HQ993481	HQ993467
Protomelas fenestratus	JF262692	JF262710	JF262746	JF262764	JF262728	JF262674	HQ993483	HQ993469
Tramitichromis intermedius	JF262693	JF262711	JF262747	-	JF262729	JF262675	-	-
Melanochromis auratus	JF262694	JF262712	JF262748	JF262765	JF262730	JF262676	-	HQ993474
Tyrannochromis macrostoma	JF262695	JF262713	JF262749	JF262766	JF262731	JF262677	HQ993489	HQ993476
Aristochromis chrystyi	JF262696	JF262714	JF262750	JF262767	JF262732	JF262678	HQ993486	HQ993472
Nimbochromis linni	JF262697	JF262715	JF262751	JF262768	JF262733	JF262679	HQ993488	HQ993475
Labidochromis gigas	JF262698	JF262716	JF262752	JF262769	JF262734	JF262680	HQ993487	HQ993473
Aulonocara baenschi	JF262699	JF262717	JF262753	JF262770	JF262735	JF262681	-	-
Hemitilapia oxyrhynchus	JF262700	JF262718	JF262754	JF262771	JF262736	JF262682	-	HQ993478
Melanochromis B&W johannii	JF262701	JF262719	JF262755	JF262772	JF262737	JF262683	-	HQ993480
Labeotropheus trewavasae	JF262702	JF262720	JF262756	JF262773	JF262738	JF262684	HQ993492	-
Lethrinops aurita	JF262703	JF262721	JF262757	JF262774	JF262739	JF262685	HQ993493	HQ993479
Copadichromis jacksoni	JF262704	JF262722	JF262758	JF262775	JF262740	JF262686	HQ993490	-
Cyanotilapia afra	JF262705	JF262723	JF262759	JF262776	JF262741	JF262687	HQ993491	HQ993477

Supplementary Table S5-2. Primers used to amplify and sequence the proximal promter regions and 3'-UTRs of several opsins from 18 Lake Malawi cichlid species.

Opsin	Primer	Sequence (5' — 3')
SWS1	Beg_F	AATGGGCAACAGAAGTGGAG
	Beg_R	CCATGTGGTCAGTGGATGAG
	Mid_F	TACCTCGCTTCTCACCCACT
	Mid_R	CTCGCATGGAGGCTAAGAAC
	F1	AAAGAGCTCAGGGTCACAATG
	R1	CCCATGAACCAGGTGAAGG
SWS2B	Mid_F	GCCTAGAGCGGTGTTCACT
	Mid_R	GATGAGACCGGAGCTTCTTG
	F3bb	TAGATTTTGATCGCAAACTCCAT
	R2b	CCAAACAGAGGTGGAAGTGC
SWS2A	Beg_F	ACCTGCTTCGAGTGGAAATG
	Beg_R	TCTGGGAACAAACACAAGCA
	Mid_F	GCAACAAAGACGCTCCCTTA
	Mid_R	AAGTGGTCCTGTGGAACCAG
	F3	GCGCGATACCTAATTTGAGC
	R2	GAACAGTGGAGGAGCTGAGG
RH2B	All_F	TGACTCACAATCAGAGATTTACCAG
	All_R	AGGACCCAAGGCGAAATAGC
	F1a	CAGTACTCCAAGGAGCTTAGCAG
	R1aa	CCACAATGTATCTCTCAACAGC
LWS	Beg_F	TCAGGCTTTCCCATAACCAA
	Beg_R	TGTCAAATCCAGACACTTTGC
	Mid_F	CGGATCTCCACTTCTCCACT
	Mid_R	TGAATCCAGCTCTAGCAAGTCA
	F0a	GGCTAACAGCTCAGGACCTC
	End_R	ACGAGACCGTTGGTGAAGAC
CNE 10a+b	F	GGCAAAGATGGAGCTGGT
	R	ACACCAGGAAAAGGATGGAC
SWS2B 3'UTR	F	CGGAGGTGACGATGAAGAGT
	R	TCGGGCTTCCTATAATAACTGC
LWS 3'UTR	F	GCTCTTTGGCAAACAAGTGG
	R	TCAAGTGCTTTTGGCAGTGT

Supplementary Table S5-3. Names, relative opsin expression results, and polymorphisms found with the proximal promoter and 3'-UTR of five opsins in 18 Lake Malawi cichlid species. Polymorphims that interrupt CRX TFBS are gray. Parts A – G cover the opsin expression results, promoters for the *LWS*, *RH2B*, *SWS2A*, *SWS2B*, and *SWS1* opsins, and 3'-UTRs for the *SWS2B* and *LWS* opsins.

Part A. Opsin Expression Results.

Species	Fish_ID	Clade	Relative Opsin Expression (%)						
			LWS	*RH2A*	*RH2B*	*SWS2A*	*SWS2B*	*SWS1*	Palette
Dimidiochromis compressiceps	2251	utaka	43.32	45.67	0.01	8.54	1.27	1.19	Long
Placidochromis johnstoni	2190	utaka	46.56	40.40	0.11	10.73	1.05	1.16	Long
Protomelas fenestratus	2333	utaka	50.20	25.49	8.13	6.02	8.11	2.05	Long
Rhamphochromis esox	2183	utaka	56.71	20.71	3.60	12.12	5.30	1.56	Long
Tramitichromis intermedius	06.437	utaka	36.73	35.80	2.39	20.29	4.62	0.17	Long
Trematocranus placodon	2188	utaka	49.81	39.99	0.23	8.28	0.19	1.50	Long
Aristochromis chrystyi	2115	utaka	4.78	43.97	25.11	1.23	17.27	7.64	Med
Aulonocara baenschi	06.830	utaka	0.42	59.18	11.66	0.23	28.15	0.49	Med
Labidochromis gigas	2358	mbuna	18.74	34.70	29.11	2.53	7.18	7.73	Med
Melanochromis auratus	2177	mbuna	1.27	41.03	33.56	0.11	23.68	0.36	Med
Nimbochromis linni	2136	utaka	2.22	34.55	35.44	0.50	15.85	11.43	Med
Tyrannochromis macrostoma	2101	utaka	29.54	46.74	6.99	2.37	4.60	9.76	Med
Copadichromis jacksoni	2453	utaka	27.72	50.34	6.73	0.00	0.20	15.02	Short
Cyanotilapia afra	2181	mbuna	0.16	51.77	36.22	0.06	0.19	11.61	Short
Hemitilapia oxyrhynchus	2171	utaka	13.25	43.92	26.12	0.03	0.03	16.65	Short
Labeotropheus trewavasae	2250	mbuna	5.89	42.64	33.25	0.07	0.68	17.47	Short
Lethrinops aurita	2338	utaka	47.06	22.24	17.89	0.29	5.48	7.03	Short
Melanochromis B&W johannii	2179	mbuna	2.16	49.51	32.21	0.04	0.54	15.55	Short
Minor Allele Frequency (MAF)	-	-	-	-	-	-	-	-	-

Continued

Supplementary Table S5-3 (continued).

Part B. CNE 10

Species	CNE 10											
	11	64	183	380	412	464	570	598	638	704	771	810
Dimidiochromis compressiceps	C	C	T	T	A	G	C	C	C	C	C	A
Placidochromis johnstoni	C	C	T	A	A	G	T	C	C	C	C	G
Protomelas fenestratus	C	C	A	A	A	G	T	C	C	C	C	A
Rhamphochromis esox	C	C	A	T	A	G	C	C	C	C	C	A
Tramitichromis intermedius	C	C	A	T	A	G	T	C	C	C	C	A
Trematocranus placodon	C	C	A	T	A	G	T	C	C	C	C	A
Aristochromis chrystyi	C	C	A	T	A	G	C	C	C	C	C	A
Aulonocara baenschi	T	C	A	T	A	G	C	T	C	G	C	A
Labidochromis gigas	C	C	A	T	M	G	C	C	Y	C	C	A
Melanochromis auratus	C	C	A	T	A	G	C	C	C	C	C	A
Nimbochromis linni	C	C	T	T	A	G	C	C	C	C	C	A
Tyrannochromis macrostoma	C	C	A	T	A	G	T	C	C	C	C	A
Copadichromis jacksoni	C	Y	A	T	A	G	T	C	C	C	C	A
Cyanotilapia afra	C	C	A	T	A	G	C	C	T	C	C	A
Hemitilapia oxyrhynchus	C	C	T	T	A	G	C	C	C	C	C	A
Labeotropheus trewavasae	C	C	A	T	A	G	C	C	C	C	C	A
Lethrinops aurita	C	C	A	T	A	G	T	C	C	C	C	A
Melanochromis B&W johannii	C	C	A	T	M	R	Y	C	Y	C	Y	A
Minor Allele Frequency (MAF)	0.056	0.028	0.222	0.111	0.056	0.028	0.417	0.056	0.111	0.056	0.028	0.056

Continued

Supplementary Table S5-3 (continued).

Part C. *LWS.*

Species	*LWS*											
	-934	-878	-773	-657	-607	-588	-442	-208	-166	-97	-70	-62
Dimidiochromis compressiceps	G	T	A	T	A	C	T	T	A	G	G	T
Placidochromis johnstoni	G	T	A	T	A	T	T	C	A	G	G	C
Protomelas fenestratus	G	T	A	T	A	C	T	T	A	G	G	C
Rhamphochromis esox	G	T	A	T	A	C	W	T	M	T	K	C
Tramitichromis intermedius	R	T	A	T	A	C	T	T	A	G	G	C
Trematocranus placodon	G	T	A	T	A	C	T	T	A	G	G	C
Aristochromis chrystyi	G	T	A	T	A	C	T	T	A	G	G	C
Aulonocara baenschi	G	G	A	T	A	C	T	T	A	G	G	C
Labidochromis gigas	G	T	A	T	M	C	T	T	A	G	G	C
Melanochromis auratus	G	T	A	W	A	C	T	T	A	G	G	C
Nimbochromis linni	G	T	A	T	A	C	T	T	A	G	G	C
Tyrannochromis macrostoma	G	T	A	T	A	C	T	T	A	G	G	C
Copadichromis jacksoni	G	T	C	T	A	C	T	C	A	G	G	C
Cyanotilapia afra	G	T	A	T	A	C	T	T	A	G	G	C
Hemitilapia oxyrhynchus	G	T	A	T	A	C	T	T	A	G	G	C
Labeotropheus trewavasae	G	T	A	T	A	C	T	T	A	G	G	C
Lethrinops aurita	G	T	A	T	A	C	T	C	A	G	G	C
Melanochromis B&W johannii	G	T	A	W	A	C	T	T	A	G	G	C
Minor Allele Frequency (MAF)	0.028	0.028	0.056	0.056	0.028	0.056	0.056	0.167	0.028	0.056	0.028	0.056

Continued

Supplementary Table S5-3 (continued).

Part D. *RH2B.*

Species	*RH2B*											
	-919	-907	-885	-860	-841	-706	-674*	-665	-352	-308	-226	-161
Dimidiochromis compressiceps	T	A	A	G	G	A	T	C	C	G	C	C
Placidochromis johnstoni	T	A	A	G	G	A	T	C	C	G	C	C
Protomelas fenestratus	T	A	A	G	G	A	T	C	C	G	C	C
Rhamphochromis esox	T	A	A	R	G	A	T	C	C	G	C	C
Tramitichromis intermedius	T	A	A	G	G	A	T	C	C	G	C	C
Trematocranus placodon	T	A	A	G	G	A	T	C	C	C	C	C
Aristochromis chrystyi	T	A	A	G	G	A	T	C	C	G	C	Y
Aulonocara baenschi	T	G	R	G	G	A	T	C	C	C	C	C
Labidochromis gigas	C	A	A	G	R	A	T	C	C	G	T	C
Melanochromis auratus	T	A	A	G	G	A	T	C	C	G	T	C
Nimbochromis linni	T	A	A	G	G	A	T	C	A	G	C	T
Tyrannochromis macrostoma	T	A	A	R	G	A	T	C	C	S	C	Y
Copadichromis jacksoni	T	A	A	G	G	C	T	T	C	G	C	C
Cyanotilapia afra	T	A	A	G	G	M	W	Y	C	G	C	C
Hemitilapia oxyrhynchus	T	A	A	G	G	A	T	C	C	G	Y	C
Labeotropheus trewavasae	T	A	A	G	G	A	T	C	C	G	C	C
Lethrinops aurita	T	A	A	G	G	A	T	C	C	G	C	C
Melanochromis B&W johannii	T	A	A	R	G	A	T	C	C	S	C	C
Minor Allele Frequency (MAF)	0.056	0.056	0.028	0.083	0.028	0.083	0.028	0.083	0.056	0.167	0.139	0.111

Continued

Supplementary Table S5-3 (continued).

Part E. *RH2B* (continued) and *SWS2A.*

Species	*RH2B*						*SWS2A*				
	-141:-137	-124	-42	-37	-29	-20	-984	-663:-657	-608:-606	-516	-493
Dimidiochromis compressiceps	del	C	A	G	A	T	G	GCATGCTA	AAG	C	C
Placidochromis johnstoni	in	C	A	G	A	T	G	GCATGCTA	AAG	C	C
Protomelas fenestratus	in	C	A	G	A	T	G	GCATGCTA	AAG	C	C
Rhamphochromis esox	in	C	A	G	A	T	G	GCATGCTA	AAG	C	C
Tramitichromis intermedius	in	C	A	G	A	T	G	GCATGCTA	AAG	C	C
Trematocranus placodon	in	C	A	G	A	T	G	GCATGCTA	AAG	A	A
Aristochromis chrystyi	in	C	A	G	A	T	G	GCATGCTA	AAG	C	C
Aulonocara baenschi	in	C	A	G	A	T	A	GCATGCTA	del	C	C
Labidochromis gigas	in	C	A	G	A	T	G	del	AAG	C	C
Melanochromis auratus	in	S	A	G	A	T	G	GCATGCTA	AAG	C	C
Nimbochromis linni	in	C	A	G	A	T	G	GCATGCTA	AAG	C	C
Tyrannochromis macrostoma	in	C	A	G	A	T	G	GCATGCTA	AAG	C	C
Copadichromis jacksoni	in	C	A	G	A	T	G	GCATGCTA	AAG	C	C
Cyanotilapia afra	in	C	T	A	G	A	G	del	AAG	C	C
Hemitilapia oxyrhynchus	in	C	A	G	A	T	R	GCATGCTA	AAG	C	C
Labeotropheus trewavasae	in	C	A	G	A	T	G	GCATGCTA	AAG	C	C
Lethrinops aurita	in	C	A	G	A	T	G	del	AAG	C	C
Melanochromis B&W johannii	in	C	A	G	A	T	G	GCATGCTA	AAG	C	C
Minor Allele Frequency (MAF)	0.056	0.028	0.056	0.056	0.056	0.056	0.083	0.167	0.056	0.056	0.056

Continued

Supplementary Table S5-3 (continued).

Part F. *SWS2A* (continued).

Species	*SWS2A*							*SWS2B*			
	-359:-349	-357	-253	-224	-217:-210	-198	-148	-611	-208	-57	-56
Dimidiochromis compressiceps	TCNGATATCTG	A	C	C	AATTAAAT	C	A	T	T	del	del
Placidochromis johnstoni	del	del	T	C	AATTAAAT	C	G	T	T	del	del
Protomelas fenestratus	TCNGATATCTG	A	T	C	AATTAAAT	C	G	T	T	T	T
Rhamphochromis esox	del	del	T	C	AATTAAAT	C	G	T	C	del	T
Tramitichromis intermedius	TCNGATATCTG	A	T	T	del	C	G	NA	NA	NA	NA
Trematocranus placodon	TCNGATATCTG	A	T	C	del	C	G	T	C	del	del
Aristochromis chrystyi	TCNGATATCTG	A	T	T	AATTAAAT/del	C	G	T	C	del	T
Aulonocara baenschi	TCNGATATCTG	A	T	C	AATTAAAT	C	G	T	T	del	del
Labidochromis gigas	TCNGATATCTG	T	T	C	AATTAAAT	Y	G	T	T	del	del
Melanochromis auratus	TCNGATATCTG	T	T	C	AATTAAAT	C	G	T	C	del	del
Nimbochromis linni	del	del	T	C	AATTAAAT	C	G	T	T	del	del
Tyrannochromis macrostoma	del	del	T	C	AATTAAAT	C	G	T	T	del	del
Copadichromis jacksoni	TCNGATATCTG	A	T	C	AATTAAAT	C	G	T	C	T	T
Cyanotilapia afra	TCNGATATCTG	T	T	C	AATTAAAT	C	G	T	Y	del	del
Hemitilapia oxyrhynchus	TCNGATATCTG	A	T	C	AATTAAAT	C	G	T	C	T	T
Labeotropheus trewavasae	TCNGATATCTG	A	T	T	AATTAAAT	C	G	T	T	del	del
Lethrinops aurita	TCNGATATCTG	T	T	C	AATTAAAT	C	G	T	C	T	T
Melanochromis B&W johannii	TCNGATATCTG	A	T	T	del	C	G	Y	T	del	del
Minor Allele Frequency (MAF)	0.222	0.167	0.056	0.222	0.194	0.028	0.056	0.029	0.441	0.235	0.353

Continued

Supplementary Table S5-3 (continued).

Part G. *SWS2B* (continued) and *SWS1*.

Species	*SWS2B*		*SWS1*									
	-55	-54	-966	-965	-952	-848	-781	-699	-539	-416	-405	-234
Dimidiochromis compressiceps	del	del	T	C	C	G	C	C	A	T	T	C
Placidochromis johnstoni	del	del	T	C	C	G	C	C	G	T	T	C
Protomelas fenestratus	T	T	del	C	C	G	C	C	G	T	T	C
Rhamphochromis esox	T	T	T	C	C	A	C	C	G	T	T	C
Tramitichromis intermedius[2]	NA	NA	T	C	C	G	T	C	G	T	Y	C
Trematocranus placodon	del	T	T	C	C	C	C	C	G	T	T	C
Aristochromis chrystyi	T	T	C	T	A	G	C	Y	G	T	T	C
Aulonocara baenschi[2]	T	T	T	C	C	G	C	C	G	Y	T	C
Labidochromis gigas	del	T	T	C	C	G	C	C	G	T	T	C
Melanochromis auratus	del	T	T	C	C	G	C	C	G	T	T	C
Nimbochromis linni	del	T	T	C	C	G	C	C	G	T	T	C
Tyrannochromis macrostoma	T	T	T	C	C	G	C	C	G	T	T	C
Copadichromis jacksoni	T	T	T	C	C	G	C	C	G	T	T	C
Cyanotilapia afra	del	T	T	C	C	G	C	C	G	T	T	C
Hemitilapia oxyrhynchus	T	T	T	C	C	G	C	C	G	T	T	C
Labeotropheus trewavasae	del	T	T	C	C	G	C	C	G	T	T	C
Lethrinops aurita	T	T	T	T	C	G	C	C	G	T	T	C
Melanochromis B&W johannii	del	T	T	C	C	G	C	C	G	T	T	T
Minor Allele Frequency (MAF)	0.471	0.118	0.056	0.111	0.056	0.056	0.056	0.028	0.056	0.028	0.028	0.056

Continued

Supplementary Table S5-3 (continued).

Part H. *SWS1* (continued), *SWS2B* 3'-UTR, and *LWS* 3'-UTR.

Species	*SWS1*							*SWS2B* 3'-UTR		*LWS* 3'-UTR
	-213	-172	-106	-83	-59	-54	-27	197	335	152
Dimidiochromis compressiceps	C	G	C	G	G	C	G	C	T	del
Placidochromis johnstoni	C	G	C	G	K	C	G	C	T	del
Protomelas fenestratus	C	G	C	G	G	C	G	M	T	del
Rhamphochromis esox	C	G	C	G	G	C	G	C	T	del
Tramitichromis intermedius	C	G	C	G	G	T	G	NA	NA	NA
Trematocranus placodon	C	G	T	G	G	C	G	A	T	del
Aristochromis chrystyi	C	G	C	G	G	T	G	C	T	del
Aulonocara baenschi	C	G	C	G	G	T	G	NA	NA	NA
Labidochromis gigas	C	G	C	G	G	C	G	C	T	del
Melanochromis auratus	Y	K	C	G	G	Y	G	C	T	NA
Nimbochromis linni	C	G	C	G	G	C	G	A	T	del
Tyrannochromis macrostoma	C	G	C	G	G	C	G	C	T	del
Copadichromis jacksoni	C	G	C	G	G	C	G	NA	NA	del
Cyanotilapia afra	C	G	C	S	G	C	G	C	Y	del
Hemitilapia oxyrhynchus	C	G	C	G	G	Y	G	C	T	NA
Labeotropheus trewavasae	C	G	C	G	G	C	G	NA	NA	T
Lethrinops aurita	C	K	C	G	G	C	G	A	T	del
Melanochromis B&W johannii	Y	T	C	G	G	C	R	C	T	NA
Minor Allele Frequency (MAF)	0.056	0.111	0.056	0.028	0.028	0.222	0.028	0.25	0.036	0.077

Supplementary Table S5-4. Complete transcription factor binding site profiles for 23 conserved non-coding elements (CNEs) in *O. niloticus* and *M. zebra*. Parts A and B cover the transcription factors AP-1, CRX, NFκB, RARα-γ, RXRα-β, and THRα-β.

Part A. AP-1, CRX, NFκB, RARα, RARβ, RARγ

CNE	bp		AP-1		CRX		NFκB		RARα		RARβ		RARγ	
	On	Mz	On	Mz	On	Mz	On	Mz	On	Mz	On	Mz	On	Mz
1	158	158	0	0	2	2	0	0	0	0	0	0	0	0
2	330	331	3	3	0	0	0	0	2	3	2	3	2	3
3	815	760	2	3	2	2	0	0	1	3	0	2	0	2
4	132	127	1	2	0	0	0	0	0	0	0	0	0	0
5	240	239	1	1	2	1	1	0	1	1	1	1	1	1
6	349	359	1	0	1	0	0	0	1	0	1	0	1	0
7	240	241	0	1	0	1	0	0	0	0	0	0	0	0
8	207	207	0	0	1	0	0	0	0	0	0	0	0	0
9	300	-	0	-	0	-	0	-	0	-	0	-	0	-
10	882	885	4	3	2	2	1	1	0	0	1	1	0	0
11	779	799	3	4	2	3	0	0	2	3	1	1	1	1
12	313	313	3	2	0	0	0	0	0	0	0	0	1	1
13	127	127	0	0	0	0	0	0	0	0	0	0	0	0
14	124	124	1	2	0	0	0	0	0	0	0	0	0	0
15	246	249	1	0	1	1	1	1	0	0	0	0	0	0
16	214	214	6	5	2	3	0	0	0	0	0	0	0	0
17	999	1404	1	1	0	1	1	1	1	1	1	1	1	1
18	428	428	2	2	3	2	1	0	0	0	0	0	1	1
19	182	191	2	2	0	0	0	0	0	0	0	0	0	0
20	311	313	3	2	1	1	0	0	0	0	0	0	0	0
21	1087	976	2	0	8	6	1	0	0	1	0	0	0	0
22	69	-	0	-	0	-	0	-	1	-	1	-	1	-
23	358	358	3	3	3	2	0	0	2	2	1	1	1	1
Total	8890	8803	39	36	30	27	6	3	11	14	9	10	10	11

Continued

Supplementary Table S5-4 (continued).

Part B. RXRα, RXRβ, THRα, and THRβ.

CNE	RXRα		RXRβ		THRα		THRβ		Total		Density	
	On	Mz	On	Mz	On	Mz	On	Mz	On	Mz	On	Mz
1	0	0	0	0	0	0	0	0	2	2	0.013	0.013
2	0	1	0	1	0	0	0	0	9	14	0.027	0.042
3	0	0	0	0	1	3	0	0	6	15	0.007	0.020
4	0	0	0	0	0	0	0	0	1	2	0.008	0.016
5	0	0	0	0	1	1	0	0	8	6	0.033	0.025
6	1	0	1	0	0	0	1	1	8	1	0.023	0.003
7	0	0	0	0	0	0	0	0	0	2	0.000	0.008
8	0	0	0	0	0	0	0	0	1	0	0.005	0.000
9	0	-	0	-	0	-	1	-	1	-	0.003	-
10	0	0	0	0	0	0	1	1	9	8	0.010	0.009
11	0	0	0	0	0	0	0	0	9	12	0.012	0.015
12	0	0	0	0	0	0	0	0	4	3	0.013	0.010
13	0	0	0	0	0	0	0	0	0	0	0.000	0.000
14	0	0	0	0	0	0	0	0	1	2	0.008	0.016
15	0	0	0	0	0	0	1	1	4	3	0.016	0.012
16	0	0	0	0	0	0	2	1	10	9	0.047	0.042
17	1	1	1	1	1	1	1	1	9	10	0.009	0.007
18	0	0	0	0	0	0	2	1	9	6	0.021	0.014
19	0	0	0	0	0	0	0	0	2	2	0.011	0.010
20	0	0	0	0	0	0	1	0	5	3	0.016	0.010
21	0	0	0	0	0	1	7	5	18	13	0.017	0.013
22	1	-	0	-	0	-	0	-	4	-	0.058	-
23	2	2	0	0	1	1	1	1	14	13	0.039	0.036
Total	5	4	2	2	4	7	18	12	134	126	0.015	0.014

Supplementary Table S5-5. Complete list of microRNA target sites identified with the 3'-UTRs of each opsin in *O. niloticus* and *M. zebra*.

microRNA	Opsin	Species	Target Seq.	Conserved[1]	Expression[2]	Ref(s)
cic-miR-1306	*RH2Aβ*	Both	TGAGGTGA	-	-	-
cic-miR-182a	*RH2Aα*	Both	TTGCCAAA	OL	E_M	a
cic-miR-182a	*RH2Aβ*	Both	TTGCCAAA	OL	E_M	a
cic-miR-219	*RH2Aα*	Both	TACAATCA	-	-	-
cic-miR-9	*SWS1*	Both	ACCAAAGT	-	E_M	a
cic-miR-96	*RH2Aα*	Both	TTGCCAAA	OL	E_M	a
cic-miR-96	*RH2Aβ*	Both	TTGCCAAA	OL	E_M	a
dre-miR-101	*RH2Aβ*	Both	ATACTGTA	-	E_F	b
dre-miR-101	*RH2B*	Mz	TTACTGTA	-	E_F	b
dre-miR-133	*RH2Aα*	Both	GGACCAAA	-	E_M	c
dre-miR-133	*RH2Aβ*	On	GGACCAAA	-	E_M	c
dre-miR-135	*RH2B*	Both	CAGCCATA	-	E_M	c
dre-miR-137	*RH2Aα*	On	TGCAATAA	-	E_M	c
dre-miR-137	*SWS2B*	On	TGCAATAA	-	E_M	c
dre-miR-139	*SWS1*	Both	CCTGTAGA	-	E_M	a
dre-miR-144	*RH2Aβ*	Both	ATACTGTA	-	E_M	c
dre-miR-144	*RH2B*	Mz	TTACTGTA	-	E_M	c
dre-miR-181	*RH2Aα*	Both	TGAATGTA	-	E_M	d
dre-miR-181	*RH2Aβ*	Both	TGAATGTG	-	E_M	d
dre-miR-181	*SWS2A*	Both	AGAATGTA	DR	E_F	d
dre-miR-190	*RH2B*	Both	ACATATCT	-	E_F	b
dre-miR-194	*SWS2B*	Mz	TTGTTACA	-	E_M	a
dre-miR-196	*RH2B*	Mz	ACTACCTA	-	E_M	c
dre-miR-214	*RH2Aα*	Both	CCTGCTGA	-	E_A	e
dre-miR-214	*RH2Aβ*	Both	ACTGCTGA	-	E_A	e
dre-miR-214	*RH2Aβ*	On	TCTGCTGA	-	E_A	e
dre-miR-217	*SWS2B*	Both	ATGCAGTA	GA	E_F	f
dre-miR-2184	*RH2Aβ*	Both	ATACTGTA	-	-	-
dre-miR-2184	*RH2B*	Mz	TTACTGTA	-	-	-
dre-miR-2189	*RH2Aα*	Both	TACAATCA	-	-	-
dre-miR-2193	*LWS*	Both	TACACATA	-	-	-
dre-miR-23	*RH2A-a*	Both	AATGTGAC	-	E_M	c
dre-miR-23	*RH2Aβ*	Both	AATGTGAC	-	E_M	c
dre-miR-23	*RH2Aβ*	On	TATGTGAA	-	E_M	c

Continued

Supplementary Table S5-5 (continued).

microRNA	Opsin	Species	Target Seq.	Conserved[1]	Expression[2]	Ref(s)
dre-miR-23	*RH2B*	Both	TATGTGAA	TR	E_M	c
dre-miR-23	*SWS2B*	Mz	AATGTGAC	-	E_M	c
dre-miR-27	*RH2Aα*	Both	TCTGTGAA	-	E_F	b
dre-miR-455b	*LWS*	Both	ACACATAA	-	E_F	b
dre-miR-461	*RH2B*	Both	CTTCCTGA	-	-	-
dre-miR-722	*LWS*	On	GCAAAAAA	TR	E_A	b
dre-miR-722	*RH2Aβ*	Mz	ACAAAAAA	-	E_A	f
dre-miR-722	*SWS1*	On	GCAAAAAT	-	E_A	f
dre-miR-725	*SWS1*	Both	TGACTGAG	GA	F_F	b
dre-miR-727	*RH2Aα*	Both	ACCTCAAA	-	F_F	b
dre-miR-727	*RH2Aβ*	Both	TCCTCAAA	-	F_F	b
dre-miR-728	*LWS*	Both	TTTAGTAA	GA TN,TR	E_F	b
dre-miR-737	*LWS*	Both	TTTTGATT	-	E_F	b
dre-miR-92	*RH2Aα*	On	ATGCAATA	-	E_F	b
dre-miR-92	*RH2Aβ*	Both	GTGCAATC	-	E_F	b
dre-miR-92	*SWS2B*	On	ATGCAATA	-	E_F	b

[1] Conserved Key:
OL: *O. latipes* (medaka)
GA: *G. aculeatus* (stickleback)
TN: *T. nigroviridis* (tetraodon)
TR: *T. rubripes* (pufferfish)
DR: *D. rerio* (zebrafish)

[2] Expression Key:
E_A: Eye (Amphibian)
E_F: Eye (Fish)
E_M: Eye (Mammal)
F_F: Fin (Fish)

[3] Reference Key:
a: Xu et al. (2007)
b: Kloosterman et al. (2006)
c: Arora et al. (2007)
d: Ryan et al. (2006)
e: Decembrini et al. (2009)
f: Wienholds et al. (2005)

Supplementary Table S5-6. Length and pairwise sequence divergence (D_{xy}) scores between *O. niloticus* and *M. zebra* for each coding and non-coding region examined.

Region	Identity (%)	D_{xy} (%)	Length	
			On (bp)	Mz (bp)
CNE_1	96.84	3.23	158	158
CNE_2	96.06	4.05	330	331
CNE_3	96.16	3.94	815	760
CNE_4	83.46	14.87	132	127
CNE_5	96.22	3.88	240	239
CNE_6	94.74	4.53	349	359
CNE_7	98.31	1.70	240	241
CNE_8	96.14	3.97	207	207
CNE_9	-	-	300	-
CNE_10	97.16	2.89	882	885
CNE_11	88.46	4.86	779	799
CNE_12	93.93	6.33	313	313
CNE_13	97.64	2.40	127	127
CNE_14	95.97	4.14	124	124
CNE_15	95.53	4.61	246	249
CNE_16	97.66	2.37	214	214
CNE_17	88.97	4.71	999	1404
CNE_18	95.32	4.84	428	428
CNE_19	91.21	9.35	182	191
CNE_20	96.14	3.96	311	313
CNE_21	93.25	7.07	1087	976
CNE_22	-	-	69	-
CNE_23	98.88	1.13	358	38
PXPRM_LWS	97.56	2.48	1000	1000
PXPRM_RH2Aα	94.80	5.38	1000	1000
PXPRM_RH2Aβ	91.77	8.60	1000	1000
PXPRM_RH2B	61.35	9.40	1000	1000
PXPRM_SWS1	71.49	26.37	1000	1000
PXPRM_SWS2A	97.19	2.87	1000	1000
PXPRM_SWS2B	81.96	16.31	1000	1000
3UTR_LWS	93.39	6.92	189	189
3UTR_RH2Aα	94.04	6.21	438	442
3UTR_RH2Aβ	93.26	7.06	465	460
3UTR_RH2B	93.15	7.18	310	319

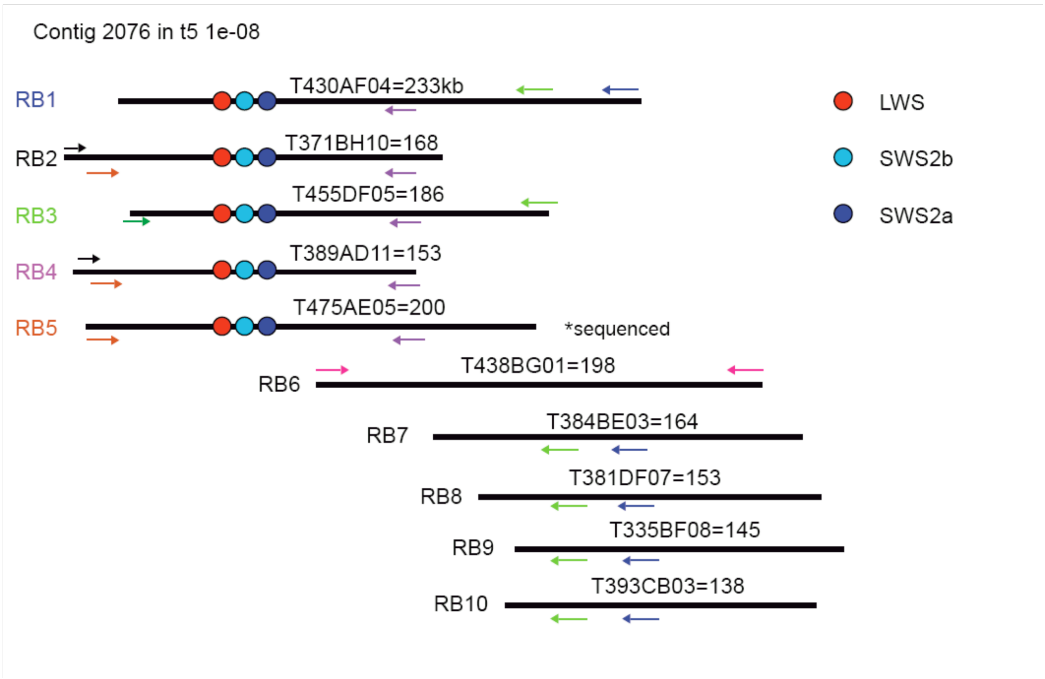
Continued

Supplementary Table S5-4 (continued).

Region	Identity (%)	D_{xy} (%)	Length	
			On (bp)	Mz (bp)
3UTR_SWS1	96.74	3.33	217	242
3UTR_SWS2A	98.37	1.64	123	123
3UTR_SWS2B	95.90	4.21	124	137
BAC_RH2	93.97	6.29	185168	48023 + 29393
BAC_SWS-LWS	95.66	4.47	171766	107685
BAC_SWS1	86.75	14.58	171942	77652
CDS_SWS1	94.54	5.67	1007	1007
CDS_SWS2A	97.73	2.31	1056	1056
CDS_SWS2B	96.69	3.39	1056	1056
CDS_RH2B	96.60	3.48	1059	1059
CDS_RH2Aα	NA	4.49	1055	1056
CDS_RH2Aβ	NA	3.78	1059	1066
CDS_LWS	96.37	3.72	1074	1074
INT-2_SWS1	89.23	12.76	60	60
INT-3_SWS1	98.63	1.48	137	145
INT-4_SWS1	95.83	3.80	108	108
INT-2_SWS2A	97.92	2.40	85	85
INT-3_SWS2A	96.96	3.22	255	255
INT-4_SWS2A	95.31	4.90	535	544
INT-2_SWS2B	87.80	15.75	113	113
INT-3_SWS2B	94.65	5.85	179	187
INT-4_SWS2B	92.46	8.08	540	542
INT-2_LWS	91.67	9.60	111	123
INT-3_LWS	88.17	13.31	84	83
INT-4_LWS	90.00	10.93	79	79
INT-5_LWS	94.38	6.79	77	77
INT-2_RH2B	84.20	17.74	1525	1405
INT-3_RH2B	87.34	14.68	239	253
INT-4_RH2B	92.99	7.31	201	234
INT-2_RH2Aβ	94.40	2.83	185	181
INT-3_RH2Aβ	93.08	4.44	119	116
INT-4_RH2Aβ	89.23	8.71	75	73
INT-2_RH2Aα	98.63	4.06	152	152
INT-3_RH2Aα	69.81	43.45	94	109
INT-4_RH2Aα	95.83	8.71	75	73

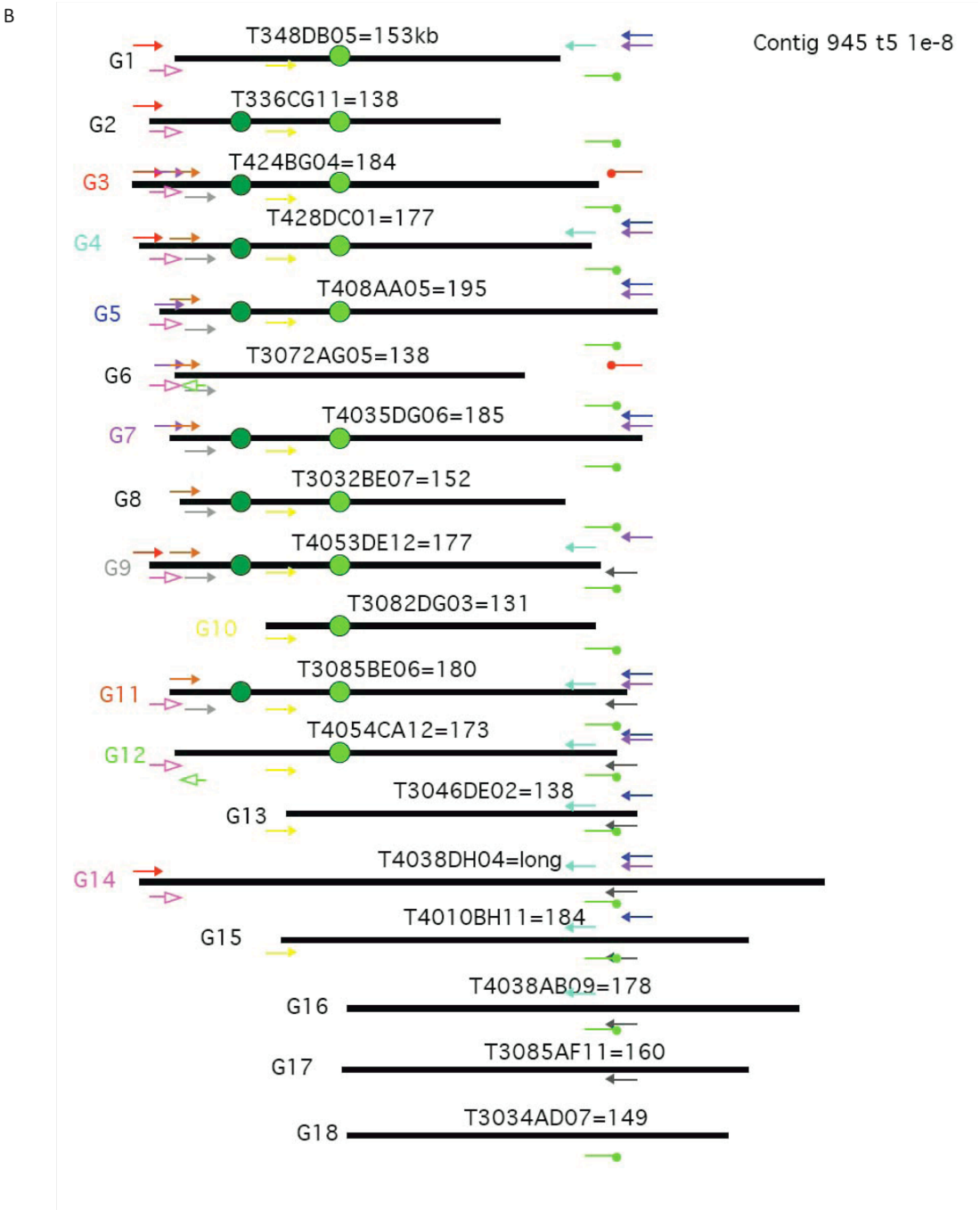
Supplementary Figure S5-1. Identification of opsin-containing BACs from Finger Printed Contigs (FPCs).

A



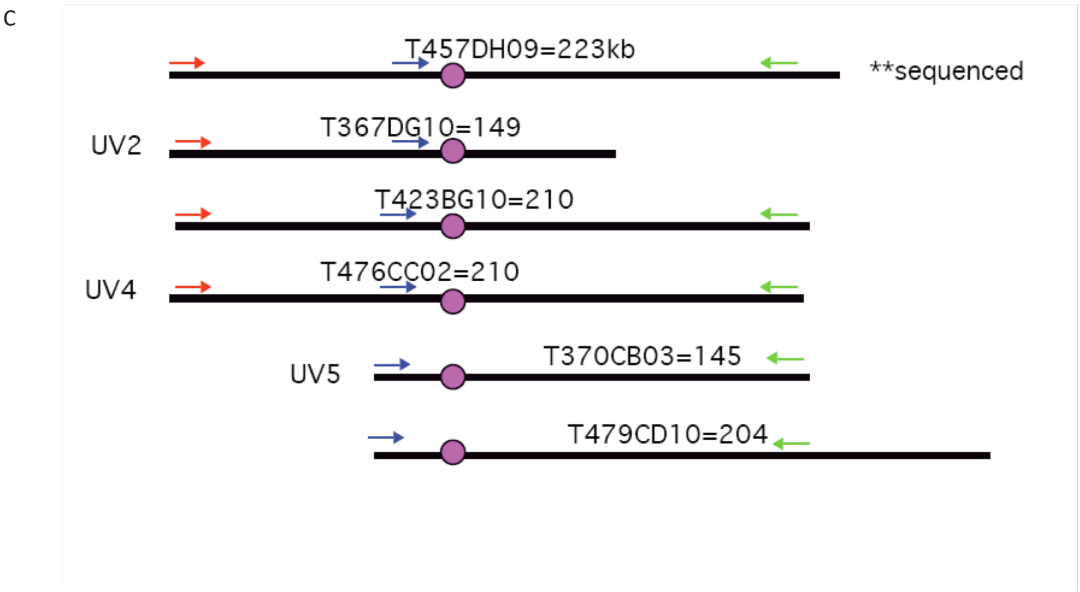
Continued

Supplementary Figure S5-1 (continued).

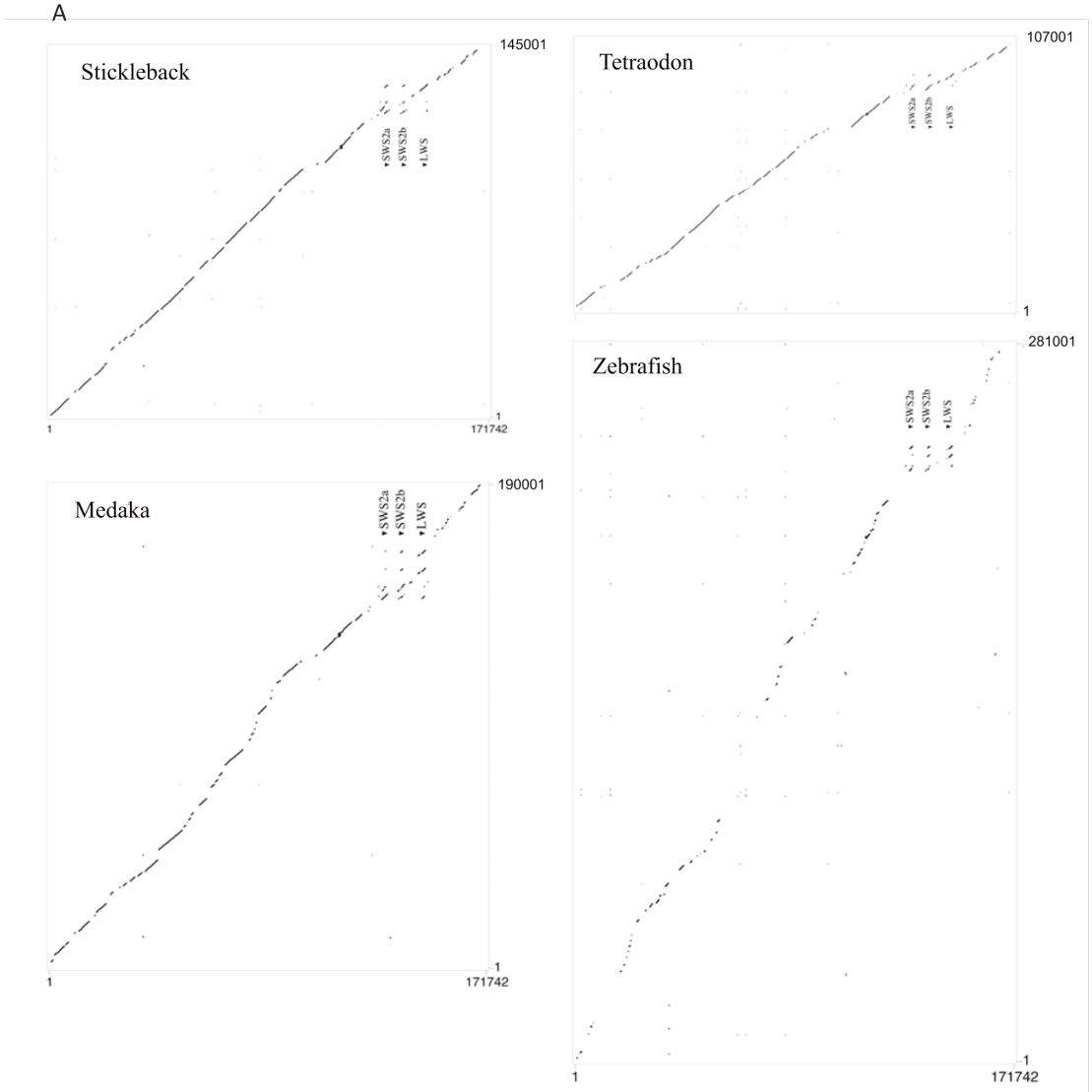


Continued

Supplementary Figure S5-1 (continued).

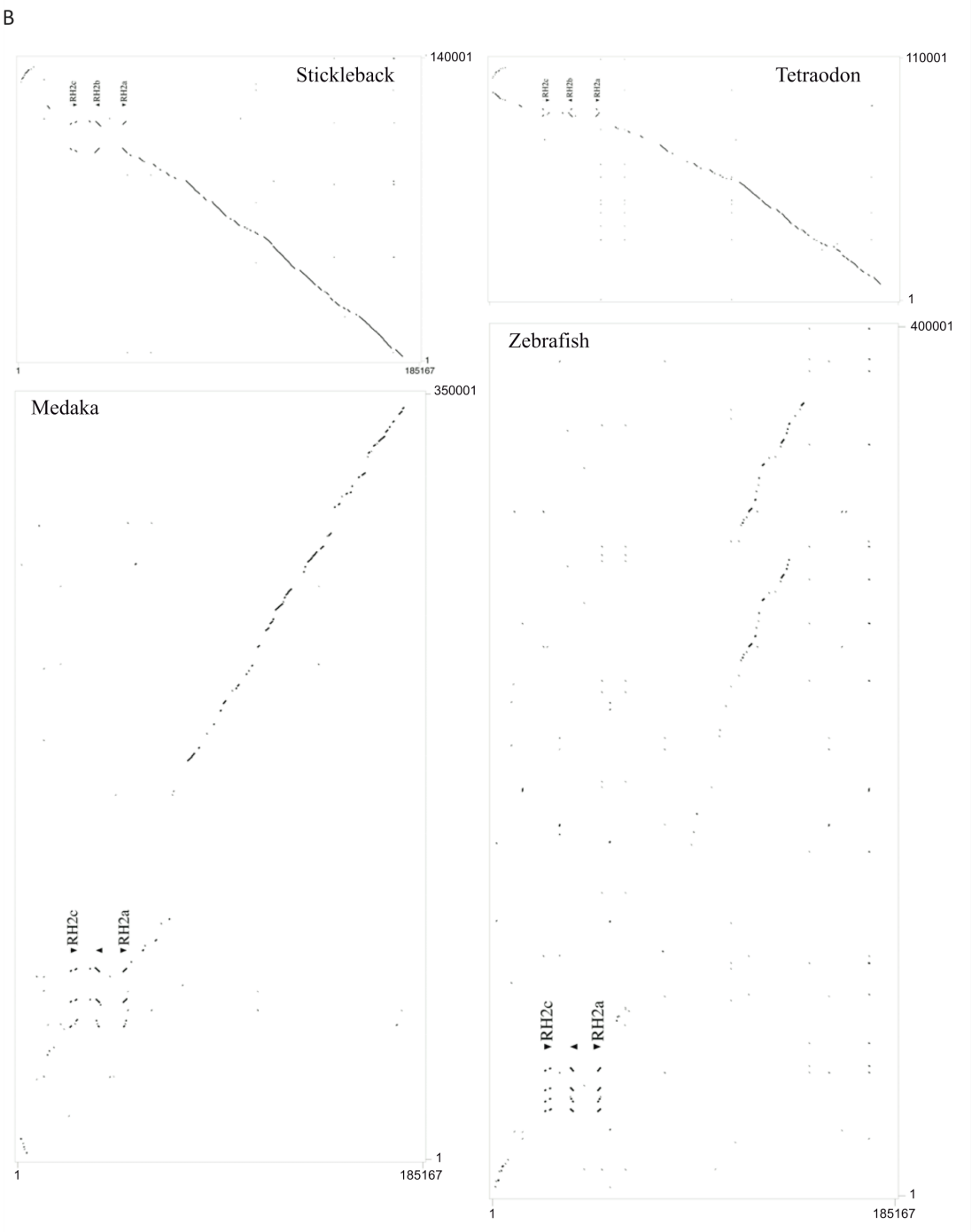


Supplementary Figure S5-2. Synteny (Pip plots) of *O. niloticus* opsin-containing BACs with the genome assemblies of five teleost species.



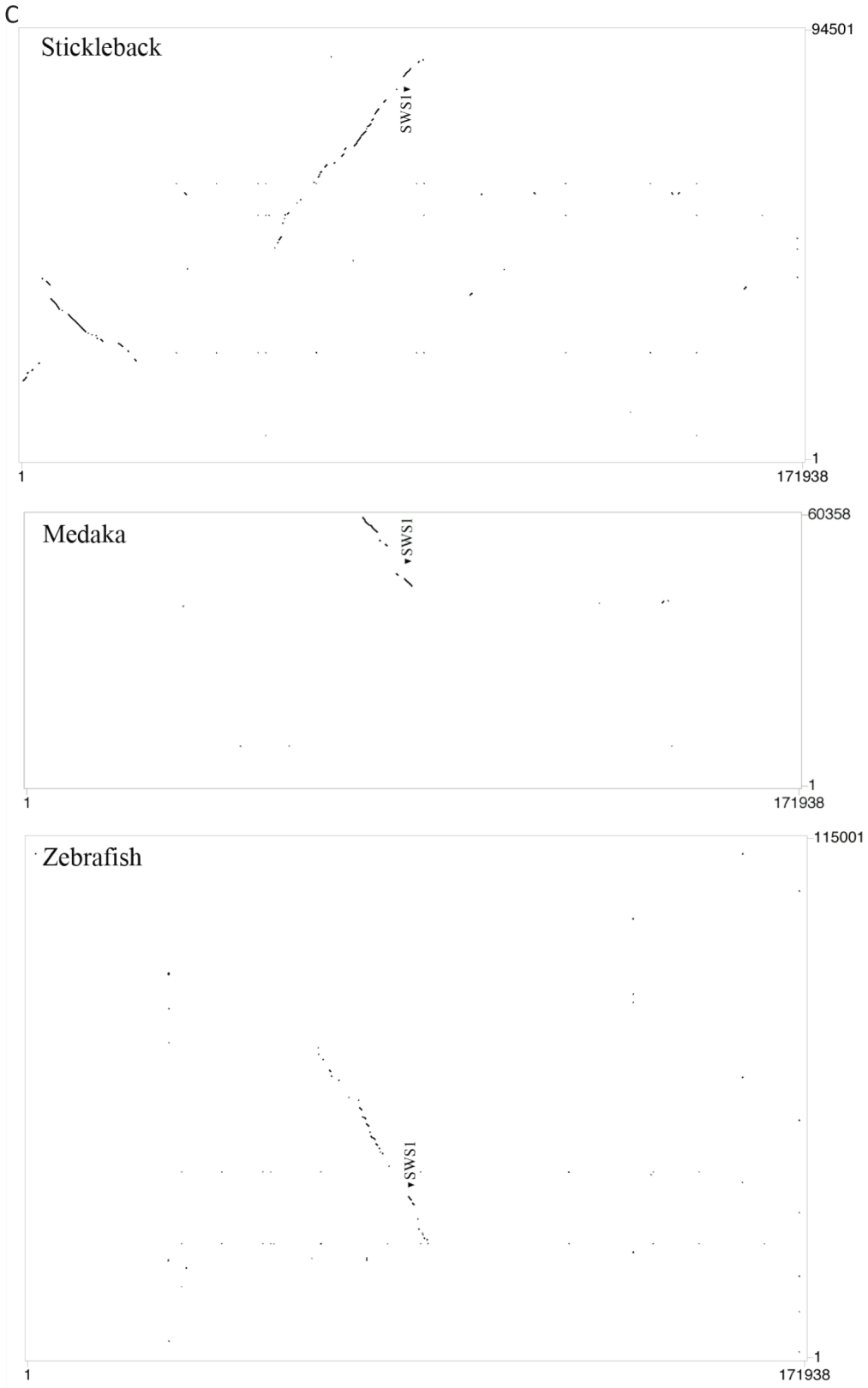
Continued

Supplementary Figure S5-2 (continued).

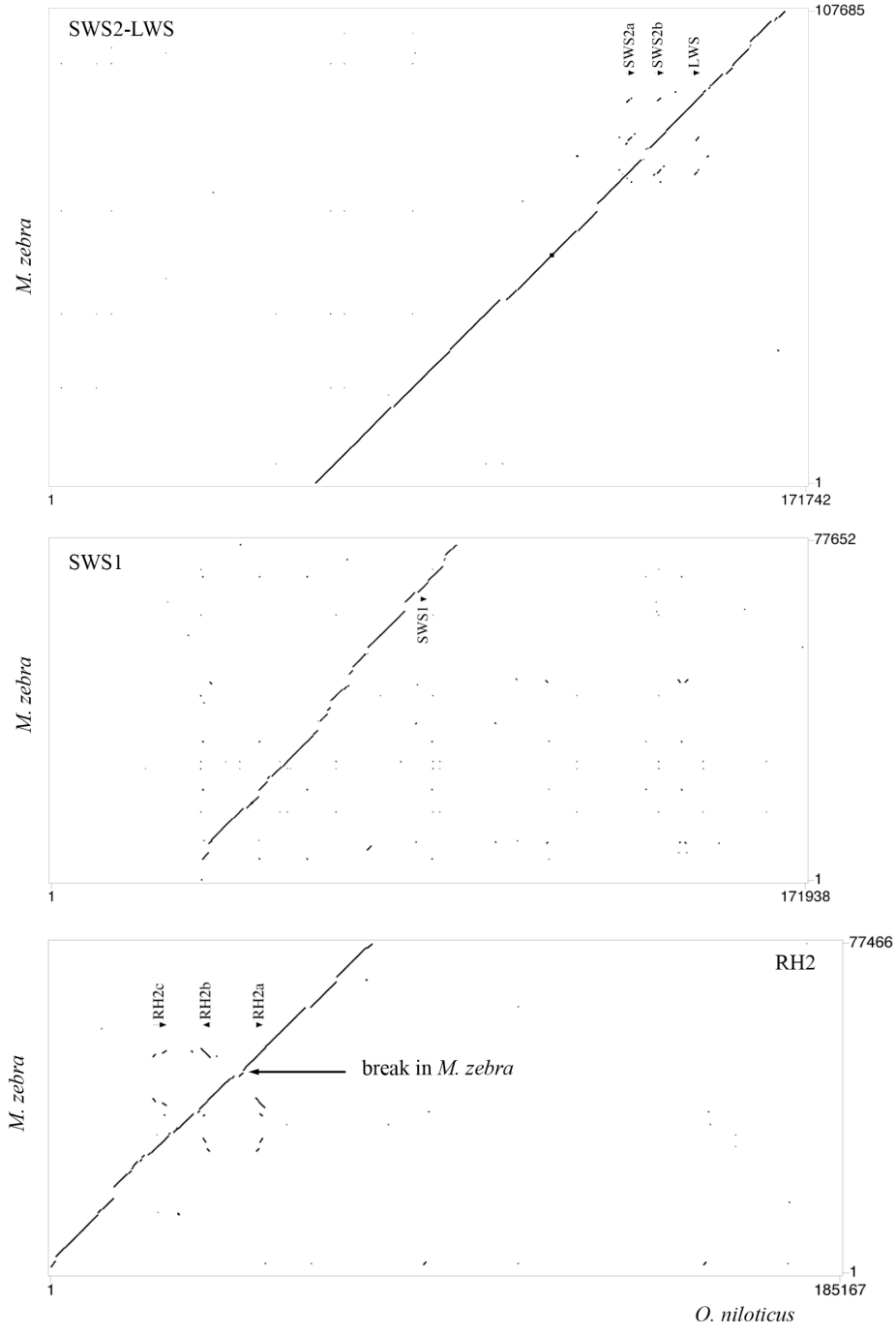


Continued

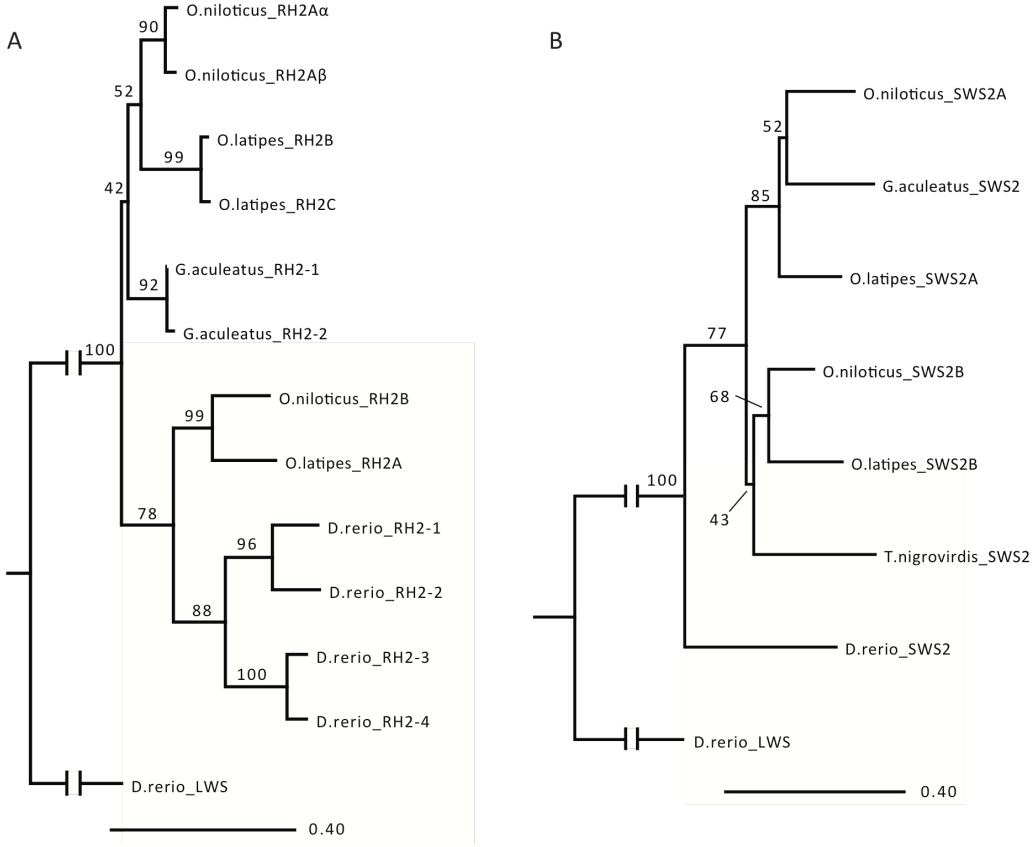
Supplementary Figure S5-2 (continued).



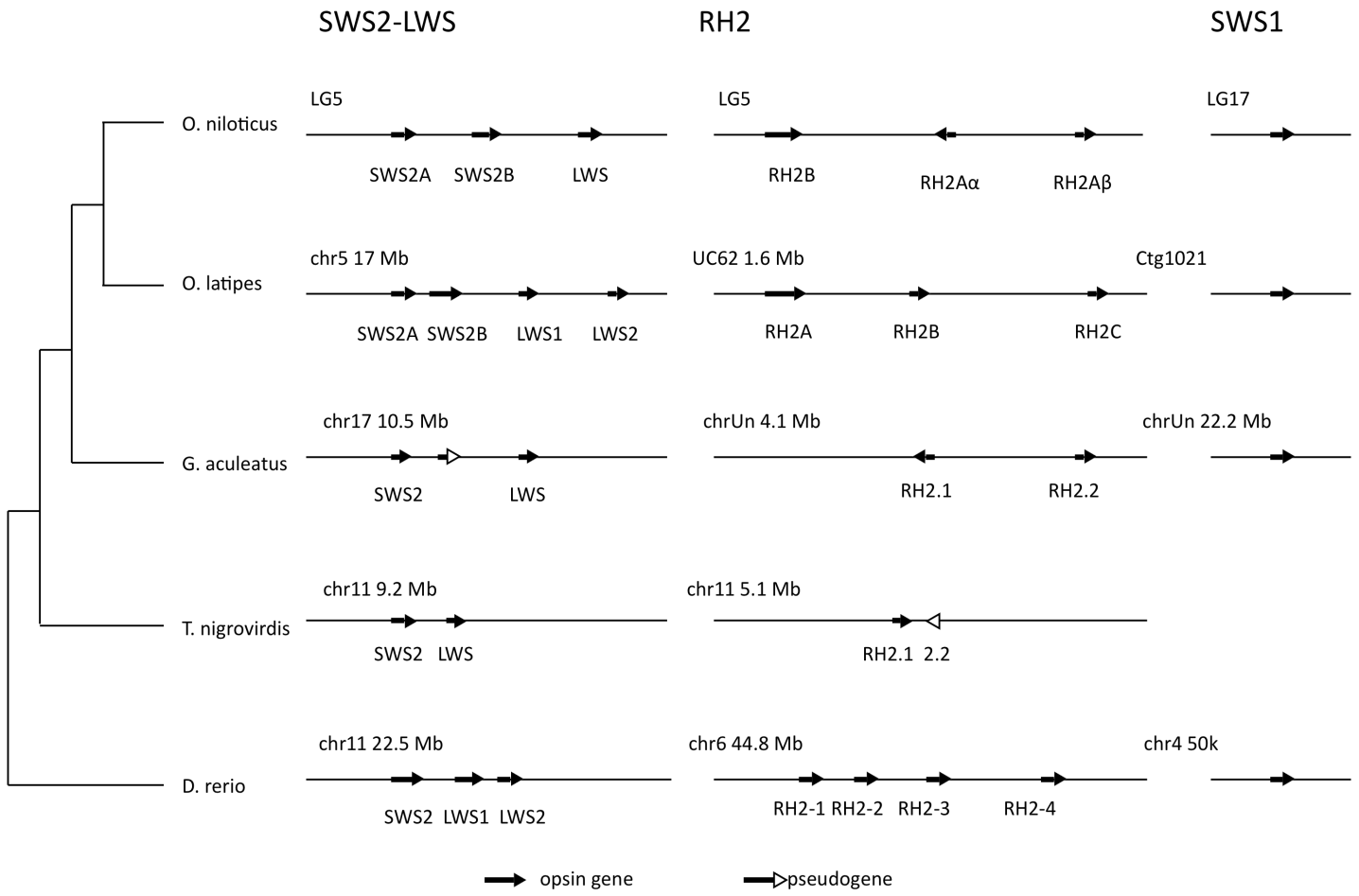
Supplementary Figure S5-3. Synteny (Pip plots) of *O. niloticus* and *M. zebra* opsin-containing BAC sequences.



Supplementary Figure S5-4. Orthology of *RH2* and *SWS2* opsin paralogs from five teleost fish genomes. (A) *RH2* phylogeny. (B) *SWS2* phylogeny. In both cases, broken lines indicate branches leading from the outgroup that were shortened to fit each tree into the figure; these do not represent missing or incomplete branch length information.



Supplementary Figure S5-5. Opsin gene content of five teleost genomes.



Supplementary File 5-1. FASTA-formatted text file of all 23 conserved non-coding elements (CNEs), opsin proximal promoter and 3'-UTR from *Oreochromis niloticus* and *Metriaclima zebra*, as well as randomly chosen *O. niloticus* non-opsin sequences from Chapter 5.

```
>Oreochromis_niloticus_CNE.1
TCTAATCTGACTGTATGACGATGTTTTGAGGATTTGGAGGATTGAGGAGGATCACCTGGTCAGGTAAATCTGAAATATCCGGATTA
CATAGGAAGTCGAACACACGGAAAAACAAAAGACTCTTATTGGATTTAGATCCGTCAGCCACCTGCTGCTGC

>Metriaclima_zebra_CNE.1
TCTAATCTGACTGTGTGACGATGTTTTAAGGATTTGGAGGATTGAGGAGGATCACCTGGTCAGGTAAATCTGAAATATCCGGATTA
CATCGGAAGTTGAGCACACGGAAAAACAAAAGACTCTTATTGGATTTAGATCCGTCAGCCACCTGCTGCTGC

>Oreochromis_niloticus_CNE.2
ACCTCACCGATATACTCCTTCAGGTCGATAAACCCGTCTCCATTCTTGTCGATGTCTTCGATCGTTTCCTGTGACAAAAAAAGAAA
GTTAAATATTTGAAACAAAACAGTGGGTCACATGACGCCCCAACATCAGTCTCACCTGCACCACGACGTCCTTCATGTACTCGTGT
TCTTCAGGATGAAGAAAAGCGGTAAATTCTTGTTTGTCTGCGATGAGGTCTCCGTTCCTGTCTGCGACCCGGAACCTCCGCTCGTC
CCTCAACATCATGTGGGTGTAGTTGTACTCTGATTCCTTCGGCGGGTCATCTGGAGCCAAGCAATGAGAGAC

>Metriaclima_zebra_CNE.2
ACCTCACCGATATACTCCTTCAGGTCGATAAACCCGTCTCCATTCTTGTCGATGTCTTCGATCGTTTCCTGTGAACAAAAAAAGAA
AGTTAAATATTTGAAACGAAACAGTGGGTCACATGACCCCACGACATCAGTCTCACCTGCACCACGACGTCCTTCATGTACTCGTG
TTCTTCAGGATGAAGAAAAGCGGTAAACTCCTGTTTGTCTGCGATGAGGTCTCCATTCCTGTCTGCGACCCGGAACCTCCGCTCGT
CCCTCGACATCATGTAGGTGTAGTTGTACTCTGGTTCCTTTGGAGGGTCATCTGGAACCAAGCAATGAGAGAC

>Oreochromis_niloticus_CNE.3
ACCCAGATAACTGCCGTATGTCACGTTCTTGTACTCTTCCCAGCTGATGCGGCCGTCGCCATTCAAGTCGAAGTCTTTCCACTGAT
GCTCCACGCTGCCAGAGATGTGCTTCCTCTGAGCGTTTTTGATCCAGGCCTTCAGCTCCTCCTCAGAGACGAAACCATCGCGGTTA
GTGTCAATTTTATCCACGATGATGCTGCAGAGACACGAATTACCAGAGCACAAAGCCAAAAACAACATCTTAATGAGAGGCAACAA
AAATAAAGACTGGACTTGAGAAATATCAAACCTAACTGTACCCAAACTGAGCTCACAGGTCCAGGAAATACACTGGACTTTATTTT
AACTGTATTGAGTTTTCTGTGAACACAGCCGAGCTGAGTATGTGAGTTGAAAATGTAAAGTTTCTCCTGCGTCTGCAGCAGCTGGA
GGTGGAGACGACCAGTGAGGACGTCTATTTCCTCACGGTCAGGATGACGCAGCTACAGGAAGTGATTATCTGAATTTTTCATTATG
CTCTGTGGACCCTGTCCCTGCACAGCCACAGTCTGCTCGGGCTGTACTTCAACGTGGAAAGCCGAGGCCGGCCTGCTTCTCTTCAT
CACTCAGAAAAACTAGACAAAGCTTCAGATGACCAGAACTCACCCGAGTCTGCGTTGGCTCTCCTCTGGAGTAAGTTGCTCGAAGG
TTTTGGCCTCCTCTTGTCCCAAAAACGCTTCGTGGTCGTAGTCGAAGTTTTTCTGGTCATCATGCTCAAGCGCGCTCAGAGGCTCC
TCATGATGGACCCGACTTTTCCTCTCTGTGGGCTTACTGCT

>Metriaclima_zebra_CNE.3
ACCCAGATAACTGCCGTATGTCACGTTCCTGTACTCGTCCCAGCTGATGCGGCCGTCGCCGTTCAAGTCGAAGTCTTTCCACTGAT
GCTCCACGCTGTCATAGATGTGCTTCCTCTGAGCATTTCTGATCCAGGCCTTCAGCTCCTCCTCAGAGACGAAACCATCGCGGTTA
GTGTCAATTTTATCCACGATGATGCTGCAGAGACACGAATTACCAGAGCGCAAAGCCAAAAACAACATCTTTATGAGAGGCAACAA
AAATAAAGACTAGACTTGAGAAATATCTAACCTAACTGAACCCAAACTGAGCTCACAGGTCCAGGAAATACACTGGACTTTATTTT
AATTGTATTGAGTTTTCTGTGAACACAGCTGAGCTGAGTATGTGAGTTGAAATTGTAAAGTTTCTCCTGCGTCTGCAGCAGCTGGA
GGTGGAGACGATTCGACCAGTGAGGACGTCTATTTCCTCACGGTCAGCATGACACAGCTACAGGAAGTGATTGTCTGAATTTCTCA
TTATGCCCGAGGCCGGCCTGCTTCTCTTCATCACTCAGAAAAACCAGGCGAAGCTTCAGATGACCTGAACTCACCCGAGTCTGCGT
TGGCTCTCCTCTGGAGGAAGTTGCTCAAAGGTTTTGGCCTCCTCTTGACCCAAAAATGCTTCGTGGTCGTAGTCGAAGTTTTTCTG
GTCGTCATGCTCAAGGGCGCTCAGAGGCTCCTCATGATGGACCCGACTTTTCCTCTCTGTGGGCTTACTGCT

>Oreochromis_niloticus_CNE.4
TGTGACACACATACAGAGACACACACACACATACGCACGCACGCAATTCCACACCCCACACACAGAAACAAAGTTCACGTTTTAAG
CTTTAAATTCCTCTTTGAATATTTTCACTCACAGTCAGCTTCTCTC

>Metriaclima_zebra_CNE.4
CCTGACACACACTGTGACACACATACAAAGACACAGACATACGCACGCGCACGCACACAGAAACAAAGTTTACGTTTTAAGCTTTA
AATTCCTCTTTGAATATTTTCACTCACAGTCAGCTTCTCTC

>Oreochromis_niloticus_CNE.5
TGTCCATGTTTCTCATACATGTGCTCACCTGGATGAGTACAGCAGACTGTGCAAGGTCAGTGGTGCTCAGATGCCAGTAATTGTTT
ACACTCATTAAGAGTCTTAGATAATGGCAGGAGAGCTGATCTTTCAAACCCTAAAGAGATCATGATGTCACTCAAGTGACATGGAT
CAAGGATTATAACAGCGGGATTTTGAGAAGGGATTTTTCTTCCTCCCACCACTGCAGCAGGCTGTCTG

>Metriaclima_zebra_CNE.5
TGTCCATGTTTCTCATACATGTGCTCACCTGGATGAGTACAGCAGACTGAGAAAGGTCAGTGGTGCTCAGATGCCAGTAATTGTTT
ACACTCATTAAGAGTCTTAGATAATGGCAGGAGAGCTGAGCTTTCAAACCCTAAAGAGATCATGATGCCACTCAAGTGACACGGAT
CAAGGATTTTGAGCGGGATTTTGAGAAGGGATTTTTTTTCCCTCCCACCACTACAGCAGGCTGTCTG
```

Continued

Supplementary File 5-1 (continued)

>Oreochromis_niloticus_CNE.6
CTGGACAGATCACTAATGTGTTGCAGGGCTAACACAGAGAGACAGACAACCATTTACCTCACTTCATGTTATTGTCCCAAGTAAAG
AATCTATGGTAAATAATTCTGGACCTCTCATATGCTAAATTTCGAGAGGATGGATTGGTGGGTTTGCATGGTAGTCTGAAGCTGGA
TGGCATGCCACCTGGTGGAGACTGATATTATTTCTTTATTGCTGGCAGAAACTCATTACAATGGCATTGCCCAAAATAAAATGCCT
AATAAGGAAATGACGGCAGTCTAGGGAGCCATTTCAAAGTAGGATACTTCTTTTTATGTTCTCAATGCTTTTATTATTTAGGGGTC
AAAAC

>Metriaclima_zebra_CNE.6
CTGGACAGATCACTAATGTGTTGCAGGGCAAACACAGAGAGACAGACAACCATTTACCTCACTTCATGTTATTGTCCCAAGTAAAG
AATCTAGGGGTAGAAGACATCTACGGCAAATAATTCTGGACCTCTCATACGCTAAATTTCGAGAGGATGAATTGGTGGGTTTGTAT
GGTAGTCTGAAGCTGGATGGCATGCCACCTGGTGGAGACTGATATTATTTCTTTATTGCTGGCAGAAGCTCATTACAATGGCATTG
CCCCAAATAAAATGCCTCATAAGGAAATGACAGCAGTCTAGGGAGCAAAGTAGTACTTTTTTTTTATGTTCTCAATGCTTTTATTA
TTTAAGGTCCAAAAC

>Oreochromis_niloticus_CNE.7
TGTCATATATGTAAAGTATTTACAGTATTAAAACAATTATATATTTCTGAAATCGAGCTGAATTTGTGAATAGTCTTTTGTTATAG
TGTTTCTATGTGGAAAAAAAATCGTCATTCACAGTTACAGGTTTGACCTGTAAGTCCACAGTCTATATTTGTAATTTTGTTGATAT
TTAATATATTTAAATTCCTGGAGAATCAATAGGAAAACACTGTTAAATTACACTGTTTTTTGTTGTTG

>Metriaclima_zebra_CNE.7
TGTCATATATGTACAGTATTTACAGCATTAAAACAATTATATATTTCTGAAATCGAGCTGAATTTGTGAATAGTCTTTCTTTTATA
GTGTTTCTATGTGGGAAAAAAATCGTCATTCACAGTTACAGGTTTGACYTGTAAGTCCACAGTCTATATTTGTAATTTTGTTGATA
TTTAATATATTTAAATTCCTGGAGARTYAATAGGAAAACACTGTTAAATTACACWGTTTTTTGTTGTTG

>Oreochromis_niloticus_CNE.8
CACCCTGAAGCCCAACAACAAGGTGACCAGGATGGTGGTGATCTTTGTGCTGGGCTTCCTGACATGATACATCCCGGCTTCGCTTT
TGTACTTGTGAACAACCACCGTCAGTCTTTTGATTTGAGATTAGCTTCTATGGTGTCTGTCTTTTCCAAGATCCCCTGCAGTCTTT
AACCTAGTTGTCTATGTTGTCCTTGATAAACAGGT

>Metriaclima_zebra_CNE.8
CACCCTGAAGCCCAACAACAAGGTGACCAGGATGGTGTTGATCTTTGTGCTGGGCTTCCTGACATGATACATCCCAGCTTTGCTTT
TGTACTTGTGAACAACCACCATCAGTCTTTTGATTTGAGATTTGCTTCTATAGTGTCTGTCTTTTCCAAGATCCCCTGCAGTCTTT
AACCTAGTTGTCCATGTTGTCCTTGATAAACAGAT

>Oreochromis_niloticus_CNE.9
ACTGATTTATTTATTTATTTAAATAAAATGCATGACTATAGATCAAAGCTAAAAAAAAAAAAAAAAGAAGCTAAAGCTAAATATATT
TTAATGTAAAGTGAGATGTTTTAGTGTTTTATTGTTACTTTTGCTAAAGCAATAGTTCTAAAACACTCCACCAGCACTCGTCATTT
CCAGCTAACTGTAAATATTTACATGTAGATGAAAACTACAGAAAGGAAGTGTGCAGCAAAATGTAAGCCACTTGTGTGACAGTGAT
GAACAGACAGTGCTGTTCTTTGTGATTTCAAAGACCCTAACT

>Metriaclima_zebra_CNE.9
NNNNN

>Oreochromis_niloticus_CNE.10
GCAGTCATTTCCTGTCACTGCAAGTTAATGACTGGAAATTACCTAAGACAGCGAATGAATAAACCAGACCTCTTAGTGACTATGGT
TATGGAGATACCAGTCGATAGTCGGGAGCAGATCTGAGTTCTGCTAGTAGTGAACTTTTGACAAACACGAATAGTTCAGAACTAGC
GGAATTCCAGACCTACCCTCGAAACATCACGCCATCACGTGGAAAAAGCAGCCAGAGTAAAACAAAACCAAAAAGCCCTTTACCTA
GCATCATCTCTACCTCCCAGGTGATATGTAGATTTACATTAAGAGATAGATCCCTAGTTAACTTCACATTTTAGAAATGTCTTTGT
TTAACCCACTGTTGCTTCCTAATCTAAGGTGGAAAATTTTAGCAAGTGCTAGCATGAAGAGAGACACAGTTGGAAATACAAAAGTG
GAAAAATATAAAGAAAATTTATAACTCCCTCCAGTTTTAGGGTCCATCCTTTTCCTGGTGTAAAGTCCTTGTCCGAGGAAGGGTTC
TTATCAAATCTGCATGTCTGCTACCTCGTGCAAGAAGTGGACTCAACAAAAGTGCAGAGTTAGGGATTAAGGTAAACCACTACAGC
AGCTTTGAGGACCCTCAGCTTTCCCCTGAAATGAACCGTTGTTCCGGAGAGATTAAAGTGTTACGGAGCGAGCATGCCAAAAGTAA
TTACTCTGTGTTAATCTGTTTTTGTGGAGGAACCTGAGTTGTCCGTATAATCTGCTTTGGCCTGCTTGGAGATGAGAAACCTTTGCT
CACTTTTCACTGTGTCAAATAGGTTAGACATTCACGGCTAAGACAGGAGAAGGTGGTGCAACACAGCTGCAGTCCTTGCAACGTCT
TCAGCTAGGAACACTGGTGCAG

Continued

Supplementary File 5-1 (continued)

>Metriaclima_zebra_CNE.10
GCAGTCATTTCCTGTCACTGCAAGTTAATGACTGGAAATTACCTAAGACAGTGAATGAATAAACCAGACCTCTTAGTGACTATGGT
CATGGAGATACCAGTCGATAGTCGGGAGCAGATCTGAGTTCTGCTAGTAGTGAACTTTTGACAAACACGAATAGTTCAGAACTAGC
GGAATTCCAGACCTACCCTCGAAACATCTCGCTGTGACGTGGGAAAAACAGCCAGAGTAAAACAAAACCAAAAAGCCCTTTACCTA
GCATCATCTCTACCTCTCAGGTGATATGTAAATTTACATTAAGAGATAGATCCCTACTTAACTTCACATTTTAGAAATGTCTTTGT
TTAACCCACTGTTGCTTCCTAATCTAAGGTGGAAAATTTTAGCAAGTGCTAACATGAAGAGAGACACAGTTGGAAATACAAAAGTA
GAAAAATATAAAGAAAATTTACAACTCCCTCCAGTTTTGGGGTCCATCCTTTTCCTGGTGTAAAGTCCTTGTCCGAGGAAGGGTTC
TTATCAAATCTGCATGTCTGCTACCTGTGCAAGAAGTGGACTCAACAAAACTGCAGAGTTAGGGATTAAGGTAAACCACTACAGCA
GCTTTGAGGACCCTTAGCTTTCCCCTCGAAATAGAACTGCTGTTCCGGAGAGATTAAAGTGTTTACGGAGCGAGCATGCCAAAAG
TAATTACTCTGTGTTAATCTGTTTTGTGGAGGAACCTGAGTTGTCCGTATAATCTGCTTTGGCCTGCTTGGAGATGAGAAATCTTC
GCTCACTTTTCACTGTGTCAAATAGGTTAGACATTCACGGCTAAGACAGGAGAAGGTGGTGCAACACAGCTGCTGTCCTTGCAACG
TCCTCAGCTAGGAACACTGGTGCAG

>Oreochromis_niloticus_CNE.11
TTATAGCAGTGAGCTAGTGTCAGATTTCCTTCTTACACACTCATGTAAAGTTAAATGTTTGCGTGACTATGGCTGAAAAAGTTTAT
AAGTTTATATATAAGTTAACATCTTTAGAGCCATCCAGGAGCCTTTCACCTCCTCCTGTCCTACATGGAGTTTTTAACTAATGTTG
GCCTGTGAGGCGAGTGGACTTGAGGAGTGCACTGATACACCAGATTAAACCATTATGCCAGCTTTGGTAAATCTCACCCCTTCCCT
CAAGCGAACTGTCAATCCGGAGAGATTAAAAGGTTCCAGAGTAAACTTGCCAGACGTGAAGGGTTACGGAGTAAACAGGTCAGACA
TCAGAGGCCAACCTAAGGCGTTCCACTAATCTGATTTGTGAGGACATGAGAAATTGTCACCTTTATTATCAGAGGCACTCACACTA
CCAGCACATCAAGAAACAGCTACAAGAGCAACAGTGTCCTCTGATGCCTTGGGTAAAATCTGAACATGCTAAAACAAATACAAATG
TAGCATTCAACATGTTATACTATATTAAACACCTGATGCAAAGTCTGGCACAGGACTGTACCCTTTAATGCTACAGATGTTGGGAG
AATACAATTTAACTCACATTAATCTGAACCAGAGGTTCCCAAAGTGTGGGGCCCGCCCCCTAGGGGGGGCGCAGAGCAATTGCAGA
GGGGGCACAGTATGAAAAGAAAAAAGAAAAGAAAAGAATGCTTGGACACTGCTAGCATAATGGACAGGTTTTTGACGGGGCAGTGG
ACACT

>Metriaclima_zebra_CNE.11
GTATAGCAGTGAGCTAGTGTCAGATTTCCTTCTTACACGCCCATGTAAAGTTAAACATTTGCGTGACTATGCCTGAAAAAGTTTAT
ACCAAGAAAAAAAAAAAAGGTTTTTCAATATCTGTCAATTACACGTGCACACAAATCTGATAAAAATAAATACTAGATGTGACATTA
AGATGAAGCTATTCACAAGAGAGTAATATTTGGTACACAGATGCAGAACAAAATCACATTTTGCTCACCTAGTTAAGGAAGTTCAC
ACTAACAGCTTGTCTACATTCACATTAAAAACAAAAAGTGGAAAAGTTAACATCTTTAGAGTCATCCAGGAGCCTTTCACCTCCTC
CTCTCCTACATGGAGTTCTTAACTAATGTTGACCTGTGAGGCAAGTGGACTTGAGGAGTGCACTGATTAAACCATTATGCCAGCTT
TGGTAAATCTCACCCTTTCCCTCAAGCGAACTGTCAATCCGGAGAGATTAAAAGGTTCCAGAGCAAACTTGCCAGACGTGAAGGGT
TACGGAGTAAACAGGTCAGACATCAGATGCCAACGTGAGGCGTTCCTGCTAATCTGATTTGTGAGGATGTGAGAAATTGTCACCTT
CATTATCACGGCACTCACACTACCAGAACATCAAGAAACAGCTGCAAGAGCAACAGTGTCCTCTGATGCCTTGGGTAAAATCTGAA
CATGCTAAAACAAATACAAATGTGGCATTTAACATGTTATACTATATTAAACACCTGATCCAAAGTCTGGCACAGGACTGTACCCT
TTAATGCTACAGATGTTGGGAGAAT

>Oreochromis_niloticus_CNE.12
GCTTTTATTTCGACATTTAACCGGATGTTGTTGTTGACGTCATCGTGTGCGCATGTCCACCTCTTTGGGATGCCGTACACGTGAAG
TCTGAGAAGAGCAAGGCCAGGGTGAAGTTATGTCCAAAGCTAGTAGGTATCCATCGGTTTTTAGTTGAAACTATTGACTCACTTAA
TACGTGACATGACATTTAAACGTCTTCGTTTTGTATTTTTTGCATTTAACAGAACAGAAACGAGCCAAACGACTCGGCTTCAGTGG
GAAGTTTAAGGTGAGTACTACAAGGCTAATGTTATCTGTCGCTAACTTATAATGC

>Metriaclima_zebra_CNE.12
GCTTTTATTTCGACATTTAACCGGATGTTGTTGTTGACGTCATCGTGTGCGCATGTCCACCTCTTTGGGAAGCAGTACACGTGAAG
TCTGAGAAGAGCAAGGCTAGGGTGAAGTTATGTCTAAAGCTAGTAGGTATCCATCAGTTTATAGCTGAAAGTATTGACTCACTGAA
TGCGTGAAATGACATTTAAACGTCTTCGTTTTGTATTTTTTGCATTTAACAGAACAGAAACGAGCCAAACGACTCGGGTTCAGTGC
GAAGTTTAAGGTGAGTGCTACAAAGCTAATGTTATCTGTCGCTAACTTTTAGCCC

>Oreochromis_niloticus_CNE.13
TTCATTAGCTTAAGGACCTTCAGGAGAAGCAGAAATTGTCTAGAGAACGAGAGCTTATGAAGAGGAGAAATTTGCAAACCTTCCAG
AATGACATCCTTCAGCGACAAAGGGAGTTTGAGCAGAGGGT

>Metriaclima_zebra_CNE.13
TACATTAGCTTCAGGACCTTCAGGAGAAGCAGAAATTGTCTCGAGAACGAGAGCTTATGAAGAGGAGAAATTTGCAAACCTTCCAG
AATGACATCCTTCAGCGACAAAGGGAGTTTGAGCAGAGGGT

>Oreochromis_niloticus_CNE.14
TCACCAAATGTGCAAGTGTTCGGATTGTATGAAGCTTGACTACCTGGAAGAAACCCATACAGGCATGTGGAGAACAGGTAAACTCC
AACCGATTTGGTTCAAACCCAGAACCTTCTTGCTGCGA

>Metriaclima_zebra_CNE.14
TCACCAAATGTGCAAGTGTTCGGATTGTATGAAGCTTGACTACCTGGAAGAAACCCATACAGGCACGTGGAGAACAGGCCAAATCC
AGCCGATTTGGTTCAAACCCAGAACCTTCTTGCTGCGA

Continued

Supplementary File 5-1 (continued)

>Oreochromis_niloticus_CNE.15
TTTGTGCAATATATCTAAAATTAGAAACATCCTGTCTTGCAGTGATGCCAAAACCTAGTTCATTCATTCATTACTTCTAGGCTGGA
CAATTGCAATTCATTAGTATCAGTTTTTTTCTGAAAGTTCCCTAGAAATCATCCTGTTAATCAAAAATACTATAACGAGTACTAAC
ATGGGGAAGAAGGAGAGATCATATTTCTTCAACACTGACTTCTCTTCATTGGCTCCCTGTTAAACCCAGAATTG

>Metriaclima_zebra_CNE.15
TTTGTGCAATATATCTAAAATTAAGAAACATCCTGTCTTGCAGTGATGCCAAAACCTAGTTCATTCATTCATTACTTCTAGGCTGG
ACAATTGTAATTCTTTAGTATCTGTTTTTTTCTGAAAGTTCCCTGGAAATCATCCTATTAGTACAAAAAATACTATAACAAGTACT
AACATGGGGAAGAAGGAGAGATCATTTTTCTTCAACACTGGCATCTCTTCATTGGCTCCCTGTTAAACCCAGAATTG

>Oreochromis_niloticus_CNE.16
TATTAGGGATTCGTCTTCAATATTTAATGAGGTAATGGACTTTTGACACACTAATGAACTAATGAAGTAAGCTTACTTTTCTCATC
AAAGACTGGACCCTCTTGTAAAGGGATTACCTCAGGTGCTCCCAAACTTTGTTCATTACACTGACTCCAGCCTGTAATCTAGTTAA
TTCAATTGGTATTAAGGTTCACATAGTGCAAAGGAGTCAAAA

>Metriaclima_zebra_CNE.16
TATTAGGGATTGGTCTTCAATATTTAATGAGGTAATGGACTTTTGACACACTAATTAACTAATGAAGTAAGCTCACTTTTCTCATC
AAAGACTGGACCCTCTTGTAAAGGGATTACCTCATGTGCTCCCAAACTTTGTTCATTACACTGACTCCAGCCTGTAATCTAGTTAA
TTCAACTGGTATTAAGGTTCACATAGTGCAAAGGAGTCAAAA

>Oreochromis_niloticus_CNE.17
TTTCTTTCCTGATAAAAGGAAGTTTTTCCTTCCCACTGTCTCAAAGTGCTTGCTCATAGGGGATTGTTTGATTGTCGGTAATTTCC
CAGAATGCTGCAAGGATCGTTGCTATAGTTGAACTGGTTATTGCGGTAAATACACATTATTATAATTTAAATGAGGTAAAAGTGCA
AAATATAGTTGGACTACTTCACTGTGTGTGAGATACTGTGATAAATAAGTAATAAGAGCTGTACATTTATGTCAGTTTTCAGGTAA
ATATAAACCATGAGTATTTACATTCCCTATCATATACAGCGCCCGCATTGTTTCCCCCCGCAATGCAGGAACTCAATATTTCACAA
CTAACTTTTTTTTTTTTTTTTTTTCTGGAAAAACTGTCACTTTTCTTATTGTATATGCTGGTAAAGGTACATTCAAAGATTTGCC
CTCTGGAGGGCGCTCCAGATCTGTATTTAAACCTTGCTGCTTCAACATTTGTGTGTTTAACTGCAGTGATGGATCTGAATGGTCCT
CAGCAGTGCTTTATAGCCACATTCATATAGCTCTGCATCTATTCTGTGTTTCACTTTTTATCTCGCTACATAAATAATGATAATAT
TGCATCCACAAAAATCCAATATAGTAGGCAACAGGCCTGCATACAGCACATGCATAGAAAATATGCACGACACAGGTTTGTAAACA
TGCCTAAAGTCTTCTACTTCTGTCCTTACCTCATTTTTAAAAAACGTTTTAACCCCCCAAAATGACACTTTTCCATTGTGCTGCAA
ATGAATCTGAAATAAAACAATGATAGATTCAATATCAGTTCATCTTCAGTTGGCAATCCTATCATTGCCATAATGCTGTTAACCAA
CCATACATGATCATGAAATTACACTGTTAATTATTTGGTGTGCACACTGCTGACCACCAGATGGCGCTTCAACCGCAAATGTAAAC
AACCCACACTAAAGGTGAAAGCATATGTGCAGTAGGTTCCCACAGAAGACATA

>Metriaclima_zebra_CNE.17
TTTCTTTCCTGATAAAAGGAAGTTTTTCCTTCCCACTGTCTCAAAGTGCTTGCTCATAGGGGATTGTTTGATTGTTGGTAATTTCC
CAGAATGATGCAGGGACCGTTGCTATAATTTAACTGGTTATTGCTGTAAATACACATTATTATAATTTAAATGAGGTAATAGCGCA
AAATATAGTTGGGCTACTTCACTGTGTGTGAGATACTGTGATAAATAAGTAATAAGAGCTGTACATTTATGGCAGTTTTCAGGTAA
ATATAAACCTTGAGTATTTACATTCCCCATCATATACAGCACCCGCGTTGTTTCCCCCCCCACAATGCTGGAACTCAATATTTCACA
AATAACTTTTATTTTTTCCCTTTTTCTGGAAAAACAGTGTCACTTTTCTTATTGTATATACTGGTAGAGGTACATTTAAAGATTTG
CCCTCTGGAGGGCGCTCCAGATCTGTATTTAAACCTTGCTGCTTCAACATTTGTGTGTTTAACTGCAGTGATAGATCTGAATGGTC
TCAGCAGAGCTTTGTAGCCACATTGATATAGCTCTGCATCTATTCTGTGTTTCTACTTTTTATCTCGCTACATAAATAATGATAAT
ATTGCATCCACAAAAGTCCAATATAATAGGCAACAGGCCTTCATACAGCACATGCTTAGAAAATATGCACCACACAGGTTTGTAAA
CATGCCTAAAGTCTTCTACTTCTGTCCTTCCCTCATTTTAAAAATAGTTTTAACTCCCCAAAATGACACTTTTCCATTGCTGCAAA
TGAATTTGAAATAAAACAATGATAGATTCAATATCAGTTCATCTTCAGCTGGCAATCCTATCATTGCCATAATGCTGTTAACCAAC
CATACATGTATTTTGTACGTGCTTTGTCTCTAGGGGCCACCGTACTGAGCATGAAATAACACTGTTAATTATTTGTAATTATTTAT
ATTTATATTTATATTTCTTCATACATTCTTATATAGTTCTATATTGTGTATTGTGTATTTTGTTGTACAGTTATTTTATTTTCAAC
TTTAATTTATATATTTTATCTTATTCTTCCCAGTTAAATTTACCCTTCATTCTAATTTGTGTTGTACAGTTATTTCATTTTTAACC
TTAATTTATATTTTATTCCTTCCTAGTTAAATTTACCCTTTTTAATTTTTCATATTTATTTCCTATCTTATTCATAGCCTTTCCTT
TTTTGTTTTCTTTAGGTCACGAGCAGTTGTCCAAGCATTTCACTACATATCGTACTGTGTATGACTGTGTACGTGACAAATAAAAT
TTGAATTTGAATTTGAATTTGGTGTGCACACTGCTGACCACCAGGTGGCGCAAATGTAAACAACCCACACTAAAGGTGAAACCATA
TGTGCAGTAGGTTCCCACAGAAGACGTA

>Oreochromis_niloticus_CNE.18
TGATGTTTCAGACCAGTGGAGCTTGTGTTCGACTGTGGTAATCACTTTTGAAGACTTTGTGAATTGAAGGAGACAGGTGTCACCTC
TAACATCCACCCTGTCTCACAGACAACCTACTCAACATCTCTGAATTACGTACAAGGTGTGAATCCAATTTGTAGGGATTGAGGAT
TAGAAAGGGACAAACCCCTCCGGTTTAACAGCTTGGTGACAATTTTTGGAAGCTAAGCGGATAAACACAAGCTCCTTGGTCTGGAA
ATCCCACTAGAGGAATTGGCATACTTATGAAAACTGGGTGACAGATGGTTTCCTGTTTATCTATCCAGAGCATCCTCATAACTTTT
AATATTGTTTCATGTAGTATCAACAGTTAAAGAAACGTGATTTGCTTGGGTGAGGATTATTATTGTGTTACTTGAAATTTAGTC

>Metriaclima_zebra_CNE.18
TGCCCTGAGAAACCAGTGGAGCTTGTATTTGACTGTGGTAATCACTTTTGAAGACTTTGTGAATTGAAGGAGACAGGTGTCACCTC
TAACATCCACCCTGACTCACAGGCAACCTACTCAACATCACTGAATTACGTACAAGGTGTGAATCCAATTTGTAGGGATTGAGGAT
TAGAAAGGGACAAACCCCTCCGGTTTAACAGCTTGGTGACAATTTTTGGTACGCTAAGCGGATAAACACAAGCTCCTTGGTCTGGA
ATCCCACTAGAGGAATTGGCATACTTATGAAAACTGGGTGACAGATGGTTTCCTGTTTTTCTATCCAGAGCATCCCCATAACTTTT
AATATTGTTTCATCTAGTATGAACAGTTAAAGAAACATGATTTGCTTGGGTGAGGTTTATTATTGTGCTACTTGAAATTTAGTC

Continued

Supplementary File 5-1 (continued)

>Oreochromis_niloticus_CNE.19
AAGCACCTTGAGGCAACTGTTATGATTTTGAGCTATATAAATAAAACTGAATTGAATTGAATTGCACCATCATGCTGCCAAATGTG
TCACTAGGGCATGTGAAAAGGCCCTTTCCACTGTGACATTGGAAGTGCTTTTCCTTCCAGTGACAGAAAAAGGTGTTTTGCTGATG
GGTTTAATGT

>Metriaclima_zebra_CNE.19
AAGCACCTTGAGGCAACTGTTGTGATTTCGAGCTATATAAATGAAACAGAATTGAATTGAATTGCACCATCATGCTGCCAAATGTG
TCACTAGGGCATGTGAAAAGGCCCTTTCCATTTTGACACTGGTTTCCTCAGCAATGCTTGTTCTTCCAGTGACAGAAAAGGTAGTT
CTGCTGATGGGTTTAGTGT

>Oreochromis_niloticus_CNE.20
TTCTGTTTTGTGGGCAGAAAGTGAAGTCTTACCCATCACATTGTCGACCATCAGCATGCACTTACCCGAATAAAAAAAATCAGAAA
GAAAGAAAAAGAAAGGCAAGTAATACATAAACCAGCCTGCTTGTTCATCAGCAAACAGATAACAGGACTGTGCACTGTTAACCTCT
AAAGGTGATGATGGCACCGAAGGGAGATTCAATGCAGTGAAAGTGGCATATGGTGTCCAAAGTACTAGACGCACATCATGACGCAC
ACACATCTCACTTCCCTAACAGTCTTCTGGCTGTAATCCTGTTACTGGGCTGC

>Metriaclima_zebra_CNE.20
TTCTGTTTTGTGGGCAGAAAGTGAAGTCTTACCCATCACCTTGTCGACCACCAACATGCACTTACCCAAATAAAAAACAATCGGCT
CAGAAAGAAAAAGAAAGACAAATAATACATAAACCAACCTGCTTGTTCATCAGCACACAGATAACAGGACTGTGCACTGTTAACCT
CTAAAGGTGATGATGGCACCGAAGGGAGATTCAATGCAGTGAAAGTGGCATATGGTATCCAAAGTACTAGACGCACATCATGACGC
ACACACATCTCACTTCCCTAACAGTCTTCTGGCTGTAATCCTGTTACTGGGCTGC

>Oreochromis_niloticus_CNE.21
AAATGTGGTTATTGTAGTTATTATAGGGTATTTGCCAAATGCATAATGACTGTGTTATTAATTAAACCAGAACTAACAAAACTAAC
TTGAAAATTATGGATTTAAAATGAAAAGCTTTGGGCTTTCTGTCTTTGTCATACCAATTTTAATTTGACATGTTTTGTTTAGAGTT
ACTAGCATTCATGTAATTACTATAGTTTTATTTATTTATTTATTTTTTTGTTAAATTCATTCTTGGGACGAACTAGGATTTGGGAA
AAAAGTTTTTTTTGCATTTTTCCATCAGTGTCCAGTCTTTTTAATTTTCTCAAATTATCTGGAAATTTCAAAAATGTAGCACAATT
AAAAATTTTAAATACAAAATTCCTAATGAGGATATATGGGATTTGTTTAAGTTCATGTAGTGGTGAACATTTACTCTCTCTACAGC
AGCAGGACTTGTCTACTTTGCAAAGATCAGTGGAATATTGTCTATGTACATTTAAATTACTTGGTTTCTTAAAAATTCAAAACTGG
AAACTGGCTGCAGTGATTTAATAAACAGAAGAAGTTTTAACATATCAAGTAAGAGTAATCAAAAGTAAAAGTACTCCTTATACATA
ACAGCTCATGTCAGAATTATATATTTTCTGCACTACATTGATGACGGAATAATAAAAGTAAAAAAGGATAAGTATAGAAAAATGTA
TTTACTTTATGTTCTATAAGTATATAATGTAGGCAGCTTGTGAAATTACTTGCAATGTATAAACATAGTACAACAGTTTTTATGCT
TTGATTGTTTATTATTATAGTAAAAGTTTATTATTCCATAGAATAGAATTTGAATGCAAATAAATCTATTATTATTACTTTATCTT
AGAAAATATATGACTATATGACATTTTCACACATTTAAAATACATTCAATTAAAAATGTTAATAATGTGTTAATAAAGTCCAATAT
ATGTTAGATATGTTGTGTGGTTCATTTTCTTCCTTTTTGCAGAGTTTTAACAGGAGTGTCAAGGAGCTACTTCAATGCTGGAAACA
CCCTGCACTCTAAAAGAACATTTTCTTGATTAAGACACTGCTCTATTCCAACCAG

>Metriaclima_zebra_CNE.21
CAATGTGGTTATGTAGTTATTATAGGGTATTTGCCAAATGCATATCTACTGTGTTATTAATTAAACCAGAACTGACCAAACTAACT
TGAAATTATGGATTTAAAATAGAAAAGCTCTGAGCTTTTTGTCTTTGCCATACCAATTTTAATTGTACATGTTTTGTTTAGAGCTA
TTAGCATTCATGTAACTAGTATAGTTTTTGTTTTTTTATTTATTTAATTTTGTTTGTTAAATTCATTCTTGGGATGAATTGGGATT
TAGGGGAAAAAAATTTACATCAGTGTCCAGCCTTTTAAATTTTCTCAAATTATCTGGAAATTTAAAAAATAGCACAATTAAGATTT
TTTAATACAAAATTCCTAATGAGGATATATGGGATTTGTTTAAGTTCATGTAGTGGTGAACATTTACTCTCTCTACAGCAGCAGGA
CTTGCCTACTTTTAACGTATCAAGTAAGAGTAATCAAAAGTAAAAGTACTCCTTATACACAACAGCTCATGTCAGAATTATATATT
TTCTGTACCACATTGGTGATGGAATAATAAAAGTAAAAAAGGACAAGTATAGAAAAATGTATTTACTTTATGTTCTATAAGTATAT
AATGTAGGCAGCTTGTCAACTTACTTGGAATGTAGAAACATAGTACACCAGTTTTTATGCTTTGATTGTTTATTATTATAATAAAA
GTTTATTCTCCTGTAGAATAGAATTTGAATGCAAATAAATATATTTTTATTAGTTTATCTTAGAAAATATATGAATAAATTTTCAA
ACATTTAGCCTAAATAAATTCAATCAAAAATGTTAATAATGTGTTAATAAAGTTCAATACATGTTAGATATGTTGTGTGGTTCATT
TTCTTTCCTTTTTGCAGAGTTTTAACAAGAGCGTCAAGGAGCTACTTCAATGCTGGAAACACCCTACACTCCAAAACAACATTTTC
TTGATTAAGACACTGCTCTATTCCAACCAG

>Oreochromis_niloticus_CNE.22
ACATTTAACCTTGAGACAGACAGGTGACCAGTTCAGGATGCGCCTTGCCTTTTCCTTTATGACGGGAAT

>Metriaclima_zebra_CNE.22
NNNNN

>Oreochromis_niloticus_CNE.23
AACCAACATTAAGTGCTACATGCCAGTATGACATCTTACGCAGGGGATTAATTACATTTAAATGTAGCCTTTCCTCCATACAAGAT
GGGCCATCTATCTCATGATGTAGCCTTTTATTTCCTGCTGATGACCTGACCCTAATTGGCATGCTGTGGGTACATTCACTGCAATT
TTCTTCAAAGAGCCTCCTCTGGACACCTGCATGTGCTCAGTCTCTTTGTCAGGGGGCTGGCAGAGGCGAGCATCACCCCTCTTTGT
TCTAATTGCTTTAATCAGATAGCGGACTGCATACAGAGGCCTTAATAACAAACTTGGGGATGCGCTGACTGTAACTCAAAGCTAAT
GTTCACTTTGAGGA

Continued

Supplementary File 5-1 (continued)

>Metriaclima_zebra_CNE.23
AACCAACATTAAGTGCTACATGCCAGTATGACATCTTACGAAGGGGATTAATTACATTTAAATGTAGCCTTTCCTCCATACAAGAT
GGGCCATCTATCTCATGATGTAGCCTTTTATTTCCTGCTGATGACCTGACCCTAATTGGCATGCTGTGGGTACATTCACTGCAATT
TTCTTCAAAGAGCCTCCTCTGGACACCTGCATGTGCTCAGTCTCTTTGTCAGGGGGCCGGCAGAGGCGAGCATCACCCCTCTTTGT
TCTAATTGCTTTAATCAGATAGCGGACTGCATACAGAGGCCTTAATAATAAACTTGGGGATGCTCTGACTGTAACTCAAAGCTAAT
GTTCACTTTGAGGA

>Oreochromis_niloticus_SWS2A_PROMOTER
TTATAAACTGCTGCTGACTGGCTGGGCATATGATAGCTGTATTTTGTTAATATTCACATACACACCAAATTGCTTAGGTTATGCAA
TTTTGTTTTTCCTGTATGAAATAGTCTATAACGTTTTTGCCCATTAAAATGATATAAAGTAAAAAAAATAAGAAAGTGTTGAGCAA
GCTAGTTTAATAATTCTATTTCCCTCCACTAACTTTCTTCCTCTTAAGTATGCTTATGCACATCCTCCTCTGTATCGGATCGTTGG
GACGACAACGAGCAAATTCTGCGATTAACCCGCCAGCTTAAATCGTACCACAGAGTGTTACCATAAAACCTTCAATCTTGCATGCT
AGTGCAACAAAGACGCTCCCTTAAGCTCAGTGAAGGCAGCAAAATGCCAAGGAGGTAACACAAAGCTTGAGTTAAATCCCCATGCA
AAATAGGGTGTAATCCACTTAACAAGAGATTATCAAATATAATGTGATTGGAAACCAGATTAGCAGGTGAAGACCGGCAGGGAACT
TCCTTAGGAAAGAGGGATTAACCTCTTTAACCAAGTTTGGATGCCAAAAGCATATCGAGGTTTGCACAGATTTTTGGAAGTCCAAA
TATGCTTGTGTTTGTTCCCAGATTCAAATACTTGTTAATTCAGATATCTGCCAAATAACCAAATCTCAGAATTTAGCATCACAGCA
CACAACACAGAACTGGAAACCACAGAAGCATTAACACCAATTAATAAAATGAAAAGCTATGAAGTCAATTAAATTAGTCAATTAAT
TACTTAGTCAATTAAATAATCAATTAACCACTTTATAGCGATCAGTGTCATGAAACCCTTTATAAACGTCAATACGACGCACACAG
CCAATAAAGACAAGAGACGATAGAATTCCTAACCGCACAGCGTTCATGTCTGCTCCTTAAATTTATTCTTTGGCAGAGCGTGTGAG
ATACCTAATTTGAGCCGTAAGATCAGCAGGAGTACGTAGTTCGAAATAAGCAAA

>Metriaclima_zebra_SWS2A_PROMOTER
TTTTTTATAAACTGCTGCTGACTGGCTGCATATCATAGCTGTATTTTGTTAATATTCACATACACACCAAATTGCTGAGGTTGTGC
AATTTTGTTTTTTCCTGTATGAAATAGTCTACAACTTTTTTTGCCCATTAAAATGAAATAAAGTTAGAAATAAGAAAGTGTTGAGCAA
GCTAGTTTAATAGTTCTATTTTCCTCCACTAACTTTCTTCCTCTTAAGTATGCTTATGCACATCCTCCTCTGTATCGGATCGTTGG
GACGACAATGAGCAAATTCTGCGATTAACCCGCCAGCTTAAATCGTACCACAGAGTGTTACCATAAAACCTTCAATCTTGCATGCT
AGTGCAACAAAGACGCTCCCTTAAGCTCAGTAAAGGCAGCAAAATGCCAAGAAGGTAACACAAAGCTTGAGTTAAATCCCCATGCA
AAATAGGGTGTAATCCACTTAACAAGAGATTATCAAATATAATGTGATTGGAAACCAGATTAGCAGGTGAAGACCGGCAGGGAACT
TCCTTAGGAAAGAGGGATTAACCTCTTTAACCAAGTTTGGATGCCAAAAACATATCTAGGTTTGCACAGATTTTTGGAAGTCCAAA
TATGCTTGTGTTTGTTCCCAGATTCAAATACTTGTTAGTTCAGATATCTGTCAAATAACCAAATCTCAGAATTTAGCATCACAGCA
CGCAACACAGAACTGGAAACCACAGAAGCATCAACACCAGTTAATAAAATGAAAAGCTATGAAGGCAATTAAATTAGTCAATTAAT
TACGTAGTCAATTAAATAATCAATTAACCACTTTATAGCGATCAGTGTCATGAAACCCTTTATAAACGTCAATACGACGCACACAG
CCAATAAAGACAAGAGACGATAGAATTCCTAACCGCACAGCGTTCATGTCTGCTCCTTAAATTTATTATTTGGCAGAGCCTGCGCG
ATACCTAATTTGAGCCGTAACATCAGCAGGAGTACGTTGTTCGAAATAAGCAAA

>Oreochromis_niloticus_SWS2B_PROMOTER
GAATTACCAGATTTTTTTTTTTTTTTTTTTACCTGATTTTTGTTTCTGGGAAAAATGAGATCCACATATTAATTTTAAATTTTAAT
AGAACAGTGGCAGTATAATGTCCTCATAAGTGGATATCAGGCCTCTGTAGATCAAAATTTGATATTTTGGTCTAGACCACCCAAAA
TGTGATGTCCATATGTGGACACCAGGTCCTAGGAGGTTAACATTTTGCCTAGAGCAGTGTTCACTGCAGTGGAGGTTTATATAAAC
TATGCTCTCTGAAGTGGTGACATAATTAAGAGGTCACAGAGGAATGTGATATCCTCGGCAGGGGAAGCAATTTGATCTTAAAGGTT
TGATAATGTTGCCAGTCAATAGAGGGAAGACCCCCAAAAAAGAGACCATGACTCATTCAAAGGACCTCTAACCACCAAATATAAAG
TAGATCTGATGAGTAGTTAGCAAGATAAGGGATGAAGAGACAGAAAACTTTTTTTTCTCTTTAATCCGACCACTTTTTCCAGAAGA
CTGACTACTCCTTCTTGGTGAAAAATAATTGTGCCAGAAAAAAAGATAAATCAACACAACCTGGAGTACTGAGGGTGAGATGCATG
CACTGACCTCTTGGAAGTCCCTGTAGATCAGATTATATGGTTCAAAAACACACTGCCCTTACACTAGGATGTAATGGGATGATCCT
TGCCAAGTCTGGTAATGCGATAGTAATCTGTGTATTAGCTATTGGGTAGGGTATGTTGTTATCGGAATGCAAGGAAAGCTTTAAAT
CCTGTCCTTTAGCCAATTATGTTGAGCCTTAGCAATCATCAAAGTTCCTATAAATGCCGCCGTTAATTCCGCTTCACACAGACGAG
CATCTCAGATATAAGACGGAAGGAAACATGAAATCTCTGAGTACTATTTAAAATTAGATGCAGCTTATTTGAAGTCTTTTAAAGTG
TTTTTATTTAATATCTTTAGATATCGATTTTGATCGCAAACACCATCTAAGAAA

>Metriaclima_zebra_SWS2B_PROMOTER
ATGCAAACTGCCAAAATGGAGTTTACAGGCAGAAAATAAAATATATACAATACACAGACTGGTGATCTGTCCCTATTGTAGTTGGG
ACAGACTCCAGTCACCACCTTCCTGCTTCCAGTGCTGGGGGGGAGGGGGGTCTGCTTGTTCTGGAACAAACAAAAATAACATCCACA
TAAACACCAAAATCCAATGGTTCAAGAAAGTTTCTGGCAAGATTATCAGTGTTACAGACTTCACCTGTCAGTGGTTTAACATTTTG
CCTAGAGCGGTGTTCACTGCAGTGGAAGTTTATATAAACTATGCTGTCTGAAGTGGTGACATAATTAAGAGGTCACAGAGGAATGT
GATATCCTCGGCAGGGGAAGAAATTTGATCTTAAAGGTTTGATAATGTTGCCAATCAATAGAGGGAAGACCCCCAAAAAGAGACCA
ATGACTCATTCAAAGGACCTGTAACCACCAACTGTAAAGTAGATCTGATGAGTAGTTAGCAAGATAAGGGATGAAGAGACAGATAG
AAAAACCTTTTCCATAAGTCTGATTACTCCTTCTTGGTGAAAAATAATTGTGCCAGAAAAAAGGAGTACTGAGGGTGAGATGCATG
CACTGACCTCTTGGAAGTCCCTGTAGATCAGATTATATGGTTCACAAACACACTGCCCTTACACTAGGATGTAATGAGATGATCCT
TGCCAAGTCTGGTAATGCGATAGTAATCTGTGTATTAGCTATTGGGTAGGGTATGTTGTTATCGGAATGCAAGGAAAGCTTTAAAT
CCTGTCCTTTAGCCAATTATGTTGAGCCTCAGCAATCATCAAAGTTCCTATAAATGCCGCTGTTAACTCTGCTTCAGACAGACGAG
CATCTCAGATATAAGACGGAGGGAAAAATGAAATCTCTGAGTACTATTTGAAATTAGATGCAGTTTATTTGAAGTCTTTTAAACTG
TTTTTTTTTTTATATTTTAGATATAGATTTTGATCGCAAACTCCATCTAAGAAA

Continued

Supplementary File 5-1 (continued)

>Oreochromis_niloticus_LWS_PROMOTER
ACTCTCACGCAGTCTTGTACCATAAAATTACTGCCAGCTTTTCAATTATCAGGCTTTCCCATAACCAAATGAAGTCTTCAATAGCC
ACCCTCACAGATCCATCATGAAGTGACTCATCTTGAATGAGCTAGTTCTAAATTTTATCATTTAATTTGTTTATTGTTTAAGTTAC
TTATTAGCCTAAAGACAGTCAATCAAACAACCAGTTTCATGAGGCTTGCTCATAATGCTCATGAGTACAAAAAGCATTCCTCCCTT
TGCTTGCCACTTGCCATAAGGGCTATGACGTGTGTTGCAGGGTCAAAGGACTGTCAAAGAATCCGATGAACCCTCTTTCCGGAGGG
ATTAAAAGTGTCTTACTTCCCTCACAGGTCAACAACTGAAACCTGCACTATTATAGGACCGCGGATCTCCACCTCCACTAAAGCTG
CTTTGCTATGCAAATCAGGGCTGACACATATACCTGTTAGGATCACAGTAGCATAAATATTAAATATTGCAAAGTGTCTGGATTTG
ACACCTCTCTGTATTCAAAAGTATAAAAAGTAAACATTTTCTTTTTAGGTTAAATGGTGATTTGAGAGATATATGCACTTTTATGT
TTCTACTTTTCATTTATACACGGCTATAGACCATTTGAGCCACTAGAGAACTCAGTGTGGAAAAGGACACAAAACTGGCATTTGTT
GGTCAAAGATCTTGAAACTAATCCTTATCTGAAGTCAGTAGACGGAAATCCACTTTCAAACAAGATTATTCTAGCATTTTGTCTTA
ACCTCATTAGTTGATTGTTTTAATTACAGTTCAGGGTTATTGCTTTGCCACTTTGCTGCGCTGCCAATAGGCACACCACAGCTATA
ATTTGATTGTGACCTAATCCAATCAAAGGTTTGTGAAGCATAGGTATAAAAGCAAAAGTCAGATCTGTAGCACCAAGCAGCCAAGG
AGGTAGCAAGTGACAAGAAAGGCTAACAGCTCAGGACCTCCTTCTAAGAAMAGA

>Metriaclima_zebra_LWS_PROMOTER
TACCATAAAATTACTACCAGCTTTTCAATTATCAGGCTTTCCCATAACCAAATGAAGTCTTCAATAGCCACCCTCACAGATCCATC
ATGAAGTGACTCATCTTGAATGAGCTAGTTCTAAATTTTATCATTTAATTTGTTTATTGTTTAAGTTACTTATTAGCCTAAAGACA
ATCATTCAAACAACCAGTTTCATGAGGCTTGCTCATAATGCTCATAAGTACAAAAAGTGTTCCTCCCTTTGCTTGCCACTTGCCAT
AAGGGCTATGACGTGTGTTGCAGGGTCAAAGGACTGTCAAAGAATCCGATGAACCCTCTATCCGGAGGGATTAAATGTGTCTTACA
TCCCTCACAGGTCAACAACTGAAACCTGCACTATTATAGGACCGCGGATCTCCACTTCTCCACTAAAGCTGCTCTGAGAGTGCTAT
GCAAATCAGGGCTGACACATATACCTGTTAGGATCACAGTAGCATAAATATTAAATATTGCAAAGTGTCTGGATTTGACACCTCTC
TGTATTCAAAAGTATAAGAAAGTAAACATTTTCTTTATAGGTTTAATGGTGATTTGAGAGATATATGCACTTTTATGTTTCTACTT
TTCATTTCTACATGGTTATAAACCAAATGAGCCACTAGAGAACTCAGTGTGGAAAAATTAGGACACAAAACTGGCATTTATTTGTT
GGTCAAAGATCTTGAAACTAATCCTTATCTGAAGTCAGTGGACGGAAATCCACTTTCAAACAAGATTATTCTAGCATTTTGTCTTA
ACCTCATTAGTTGATTGTTTTAATTACAGTTCAGGGTTATTGCTTTGCCACTGTGCTGCACTGCCAATAGGCACATCACAGCTATA
ATTTGATTGTGACCTAATCCAATCATGGGTTTGTGAAGCATAGGTATAAAAGCAAAAGTCAGATCTGTAGCACCAAGCAGCCAAGG
AGGTAGCAAGTGACAAGAAAGGCTAACAGCTCAGGACCTCCTTCTAAGAACAGA

>Oreochromis_niloticus_SWS1_PROMOTER
GGTCCAAAATGCCGCAGCACGTCTTTTCACCGGCACTAGAAGGTATTATTCCATCACCCCTGTTCTAGCCAACTTACACTGGCTTG
ATATCAAATATCAAATCAATTTAAAAATCTTATTGTTTACGTATAAAATCCTGACATTCGCGCATCAGATTCTATTCAATCTTTTA
AATCTAAAAACGTACTTGTTTAACCTGGTTTTTAAGGCTAATTAGTGTGGTTTTTATCTGGATCTGCATTTATATTATGTAATTAT
TTTATATCTGTAATTTATATTTTATATTGTGATTTTATCTGTATCATGGTTTTACTGGAAAGCACTTTGATGCACTTTGGTCTTTA
AATGTTCTATACAAATAAATTTGGTATGATATGATATGATATGAAGTGTACCCCGCCTGTCACCCACCGACCCGCCCACAACCCTC
AGCTGGTTGGATGGATGCTGATGTTTCTAGTTATTATGTGTAACCAAACTTAAATAGTACAAATAAATCATTTAGCATAACTAAAG
TTTGAATTTCATGTTATTAGCTGTTAATATGATTGCTGTTTTATGTTTAGATCACTAATAATTTAGTTAGTGAAATCAATCATTAA
ATTAAGTCTGTTTGGTAACATCAACTGACCACAGGGTTCCAGGTTCAATTCCTCCGAGTTAAAGGGCTAATTCCATAGCCATTTAC
TAAAACAATAAATAAATAAATAAATAAAACAAGTCAGAGTTTTATAAAAACAATGTATAGGTGAAAGAAATGAAGACGTGCATGCA
ACCACGATCACTTTTAAAGCAGGAAGTATGTTTTGAGTTTTCTGAAGTGAACTCGGGTGGATCCGTGTGTTCAAAGAGAAAACCTT
CTGTAAGCAGATTAAGGAGTCAGAAGTTCTTAATCCTGAAAGTTTAGAAAAATCCCAGCAGCATAATCTTTGCTGTAAGTGGTTTA
TGAGCGTATATAAGAGGCTGACACAGCGGCAGCAGCAGAGAGCTCAGGGTCACA

>Metriaclima_zebra_SWS1_PROMOTER
CTTCCTCACGCTGAACCCCTTTAACCGTTTCAGTGGTCGTGAGTCTTCTAATCTGACTGTGTGACGATGTTTTAAGGATTTGGAGG
ATTGAGGAGGATCACCTGGTCAGGTAAATCTGAAATATCCGGATTACATCGGAAGTTGAGCACACGGAAAAACAAAAGACTCTTAT
TGGATTTAGATCCGTCAGCCACCTGCTGCTGCTCTTCATCATCAGGCGTCTTCATCGCCCTGCAGTGGGCCTGACAACAGCTTGTG
TTTATTACACTAAAAACTTTATAAACCCATCACAAACCATATCACACAGCAGGGACTTACCTCTTCATCTGTAAGAAGGATTTTTA
GAGTTGGCAGCAGAGCAACAGTCAGCTCTGTTGCCTCACTAAAAGAGATCTTTGTTTGAATCTGTGACCTGTCCAAGTGTACCTCG
CTTCTCACCCACTGACCTCTCCACAACAGTGAGCTGGTTGGCGGGATGCTAATGTTTCTAGTTATTACGTGTAACCAAACTTAAAG
AGTACAGATAAATCATTTAGCATAATTAAAGTTTTACTGTCATGTTATTGGCTGTTAATATGATTGCTGTTGTAAGTATGTGTTGA
TCACTAACAATTTAATTAATTAAATCAATCATTAAATTAAGTTTGTTTGGAAAAAGAGGGAAAACTCATCCACTGACCACATGGTT
CTAGGTTCAATTCCTTGGAGTTAAAGGGCTAATCCCAGAGCCATTTACCAAAATAATAAATAAATATTTAAATAAGACGTGCATGC
GACTGCGGTCACCTTTAAAGCACAAAGTTTTTTTTTGAGCAGTGAGGTGAACTCGGGTGGATCTGTGTGTTCACAGAGAAAACCTT
CTGTAAGCAGATTAAGGAGTCAGAAGTTCTTAATCCTGAAAGTTTAGAAAAATCCCAGCAGCATAATCTTTGCTGTAAGTGGTTTA
CGAGCGTATATAAGAGGCTGACACAGCGGCAGCGGCAAAGAGCTCAGGGTCACA

Continued

Supplementary File 5-1 (continued)

>Oreochromis_niloticus_RH2B_PROMOTER
TTCTTTTTGTACAAAAAAAGTTTAAAAATTACTTTGTATTTGCTTTCTGATGACCAGTGAGTGCATTATGAAAGAAGGTGGAAAAG
GAAGCAACTAGACTAGGGCAGACTAATGGGACTAATCAACATGGCACTGTGACGTGTGGGCGTACCCATTCTATTTCCAATTTCAG
GACACAGGCACAAGTTCAGCATTGGTAAATCACTGAGGTTTGTGGGTCACTGTATCAAAACACCGCCTGATATAGTGTTTCAAAAA
GTGCTGATGACTTGAAAAGGAGTTGTGAAAAACATATTTTATAAATAAACTGCAAATATCACCTGCCAAGGTTTGCAACACTCTTG
TGTGTGGACCAGGTTGAGTCTGCCCACTAAGAAATTGTGTGGAAAAACTTTGTGGACACGAACTTGGAAATTTTATATAAATGTCA
CTTACTGCTAAACTCCCACATTGGAGTTGCACTCCAATTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT
TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCCATCATTTATTTAAAATATATAAACTC
CTGAAGTCTGAAGCTTGGCCTATGCTAAAGTTGTGACAGAAATTTGAGATCGCCAAGAAATTCTGTCTTGTTTATGAAAGCAATGG
TCTACGAGTAAAACGCTCCCAGTTTAAATGGGAGAATAACAGGTAAAAGCCTACTCTCAATGACCTAGTGACCATCATTTGGTAAC
AGTCATGGTTTTAATTACTAAATGCTTCCTAAGAAAATGTGGAGAAAAAAAAGGATCACCAAGTAACAAGTGATTATACTCAGAGA
TTATGGGCTGAATTCTCAGATCCTCTTAGCTCATCTTCATCATCCCCTGGGGGCTATATAACTGTGGCTGCTCAGTGGTGTTTAAG
AAGTTAGAAAGCAGACTCTACTGCATTAACAGTACTCCAAGGAGCTTAGCAGAG

>Metriaclima_zebra_RH2B_PROMOTER
GACTCACAATCAGAGATTTACCAGACTTATCATACCCAAAACTATATAATTACACAAATTAAAATCACTTGTTTTTTCACTAACAC
TTTACAATACAGCCCATGACTAAGCTCAATGAGGAGAGGACTGTAAGGGTTTCAAGTCCTGTGGTGTTAATGGGAATGTGTCCCGT
GAAAAGAAAGAAAGAAGAATGCACAAAGCATCTATTTATAGCATTTACAGTATGTTGGCATTGTTGGCCTTTTACAGCATGTCCTT
TTTTCTTCTTCTGGTCTTTTTGTAGTGGAAAAGGAAGTGAAAGACTAGGGCAGACTAATGGGACTAATCAACACGGCGCTGCGACG
TGTGGGCGTATCCATTCTATTTCCAATTTCAGGACACAGGCACAAGTTCAGCGATGATAAATCACTGAGGTTTGTGGGTCTCTGTA
TCAAAACACCACCTGATATAGCGGTGCAAAAAGTGCTGGTGACTTGAAAAACAGTTTATTTTTAGTTTTGTTTCATTTTATAAATA
AACTGCAAATATCACCTGCCAAGGTTTGCGACACTCTTGCATGTGGACAGGGTTGAGTCTGCCCACTAAGAAATTGTGTGGAAACC
TTTGTGAACACGAAGTCTGAAGCTTGGCCTATGCTAAAGTTGTGACAGAAATTTGAGACCGCCAAGAATTTCTGTCTTGTTTATGA
AACCCAATGGTCTATGAGTAAAACGCTCCCAGTTTAAATGGATGAATAACAGGTAAAAGCCTACTCTCAATGACCTAGTGACCATC
ATTTGGTAAAAGTCATTGGTTTTATTTACTAAATGCTTCCTAAGAAAATGTGGAGGAAATAGGATCGCCAAGTAACAAGTGATTAT
ACTCAGAGATTATGGGCTGAATTCTCAGATCCTCTTAGCCCATCTTCATCATCCCCTTGGGGCTATATAACTGTGGCTGCTCAGAC
TCTGCTGCATTAACAGTGCTCCAAGAAGCTTAGCTGGACAACTGATCAGCAGAG

>Oreochromis_niloticus_RH2A-alpha_PROMOTER
CTTTGCTAGTCAGCTGTTTAGTTGAGTCTGCGTAAAGAGTCTACTTGTGTGGATGGTGACGAGGTTTCAAGTTACACCCTGCGATT
TTTATACCTCTCCACTAAACCTGTCAAACAGCCAAAATGGGATTAGAGGATTGTGAGGTAGTGCAAAAGCTAAGAGGTCACTAATT
AATCTTAACACCGGATAGACAGGACAGGGTCCAAACTAATTTGCTCCACTTTCCAAGTCCGGAGGAATTCTCCTAATCCTTACTCG
AGAAAGGGAATTACCTTCCCTCAGCTTTTAAAGTCTGATGAGACATTAATCATGAGCAAACACACAAATCATACCTGAAATATGAG
AACTGTCTGTGTATATAAGAACAAGTGTATTGATTGTAGAGTGTTAAATATATTATCTACAGCAGATAATGTTTAGTGATAGTGTG
CTTTACATAGTCTGTGTCTTTTAATGTGAGCTATGTATGTGTTGAAACAAGTGTTTTACCTGCTTCATCTGCCTTCTTCTAATCCA
ATTCACTGTGGGTGCTTTCTTTCCTAGAGTGCTTAACTACCCTACATTAAATTTCTTATGTTTTAATTATTTCAGACCAGACCTAC
AGTCAAGCATTTTTCACCCTAACAGTCTGTCTCCATATGGAGGAACTTTGGGTGTCATTATCCTGATCGGATTAACCTCAGTATCT
CACCTGTGGTAGCCTAGCTGTGATGTGGTAAAAGCTTACTGTCACAGGTCGGAAAGAAAATGTCCGGTAAAGACTTAGCCAATTTG
TGGTCTTCTTTAAGCTACTGCAACACTATCAGTGGCCAAGAACATCTAATATCAGTTATCCTTCTGTTTATTTGACAGATAGAAGT
GATAGATGATGATTGGGCTACATGATGGAGAAACCACATCATTTCTCTGTTACCTGTGATTAATTCAGATCAACTTGCTTGTGATG
TTCTGTTTAGTCCCAGTTATGTGATTTGATATGATAATAAATTTAGTTTCTTTT

>Metriaclima_zebra_RH2A-alpha_PROMOTER
CTTTGCTAGTCAGCTGTTTAGTTGTGTCTGTGTAAAAAGTCTACTTGTGTGGATGGTGACGATGTYTCAAGTTACACCCTGTGATT
TTTATACCTCTCCACTAACCCTGTCAAACAGCCAAAATGGGATTAGCGGGTTGTGAGGTAGTGCAGAAGCTAAGAGGTCACTAATT
AATCTTAACACCGGATAGACAGGACAGGGTCCAAACTAATTTGCTCCCCTTTCCAAGTCCGGAGGAATTCTCTTAATCCTTACTCA
AGAAAGGGAATTACGTTCCCTCAGCTTTTAAAGTCTGAAGAGAAGTTAATCATGAGAAAACACACAAATCATACCTGAAATATGAG
AAATGTCTGTGTATATAAAAACAATAACGTGTACTGATTTTAGAGTGTTAAATATATTAACTACAGCAGATAATGTTTAGTGATAG
TGTGCTTTACACAGTCTGTGTCTTTTAATGTGAGCTATGTATGTGTTGAAATGTATGTGTTGCCTTCTTCTAATCCAATTCACTGT
GGGTGCTTTCTTTCCTACAGTGCTTAACTACCCTACATTAAAAATTTTATGTGTTTATTAATTCAGACAAGACCTACAGTCAAGCT
TTTTTTACCCTCACAGTCTGTCTCCATATGGAGGAACTTTGGGTGTCATTATCCTGATTGGATTAACCTCAGTATCTCACCTGTGG
TAGCTGTGGTGTGGTAAAAGCTTACTGTCACAGGTCGGAAAGAAAATGTCCGGTAAAGACTTAGCCAATTTGTTGTCTTCTTTAAG
CTACTGCAACACTATCAGTGGCAAAGAACAGCTAATATCAGTTATCCTTCTGTTTATTTGACAGATAGAAAGGATAGATTATGATT
GGGCTACATGATGGAGAAACCACGTCATTTCTCTGTTACCTGTGATTAATTCAGATCAACTTGCTTGTGATGCTCTGTTTACTCCC
AGTTAAATGTGATTTGATATGATAATAAATAAATTTACTTTCTTTTATTATTTT

Continued

Supplementary File 5-1 (continued)

>Oreochromis_niloticus_RH2A-beta_PROMOTER
GTCTATTCTAAGCTGTTAACTGTTGGCGTGAGTTCACATGTTCATGGTCAAATGTTAAACGTCTCTTTGAGGCGTTAATGAGGCAT
TGGACTAGTTGGCTGAATCAGCTACTGTAACACCTACTGGAATAACAAAAATATAGCAATTAGTTCTTCAGTTGTGTTGTTTTGTT
TTATGACGTTTTGAATACAGTCAGTTTTTTTTTCCTTATCTTTGTGTTTGACCCCACCACGCTGATGGGGTACTGAACACGCCACA
TGTTTTAAAGTAAATAAAAGATAAATCACTATATCAACAGTCTTGCTTTACCACGGATCATTGTAAATAGTCATTGGTATTCATAT
TTTTAATGGTCACAATTTCCTTAATCAGTTTGGTTCCTTTTGATCATTTGATAAACTGTTTTTTAAAACTATTTTTATGCATATCA
GTTCATTTTTCTGTCTCTTAATTGCAAAAACTTCGTTCAAATGCTCCAAATCACATATTTTAATACAATAGGTTTAGGATTGTGAT
GAGGTTATGGTTACTTTCGTTCTGAGGTTTCATTTTTAAATACTCCAACCATAACACTTATTTATTTATTTATTTTACTTTAAACT
GAACTTTTAAACCTATTTGAAACATTTTTTGTACTACACTGAGAACTTAAGCAATGACAACGAAAAGCAAAGAAAAATATTGTTTT
CCTATAAGAAAGGGGTGTGTAGTGTAAAGCAGCATTTTCTGCAAAATTAGGAAATTTATCTCAAGAGGCCAAGATATTTGACCCTG
AGGTCATTGTCCCTCTTTGCTTTTTAAAGTCAATTATCAGTCAGACTTGATCTGTTCACTTCCTCACAATCCTCTTGTAATCTAAT
TTTGCTAATTTGAGGATAACCCTGAGCAGGTATAAAACACCAAGGCAGCCATAAACTACCTGTGTGTGGTTGGTTTACACCTCTCT
AACACTACTGGGATATTCCATCAGCTGAAACTCAAGCAAACTAGACAACAGAAG

>Metriaclima_zebra_RH2A-beta_PROMOTER
TTAACTGCTGGCGTGAGTTCACATATTCATGGTCAAATGTTAAACGTCTCTTTGAGGCGTTAATTAGGCATTGGACTAGTTGGCTG
AATCACCTACTGTAACACCTACTGAAATAACAAAACTATAGCAATTAGGTCTTCAGTTGTGTTGTTTTGAATACACTCTAGTTTTT
TCCTAATCTTTGTGTTTGACCCCACCACGCTGATGGGGTACTGAACACGCAACATGTATTGAAACAAATAAGAGATACATCGCTAC
ATCAACAGTCTTGCTTTACCATGAATAATTGTCATAGGTATTCATATATTTAATAGTCACAATTTCCTTAATCACTTCAGTTCCTT
TTAGTTTACTTTGATAAAGTTCCTTTGATAAACTAGTTTAAACAAATGATACAAAAGTTTAAAAACACTTCATTTTTCTATCTCCT
AATTGCAAAGACTTGGGAGAAATTCTTTCAAATGCTCCAAATCACATATTTTAATACAATAGTTTTAGGATTGTGTTAAGGTTATG
GTTACTTCCTTTCTGAGGTTTCTATTTTTAAATACTCCAACCATAAAACCTATTTACTTATTTATTTTACTTTAAACAGAATTTCT
AAAACTATTTGAAACAATGTTTGTACTATACTGTGAATTTAAGCAATGACAACCAAAAGACAAAAGAAAAATATCATTTCCTACAT
TATAAGAAGGGGTGTGTAATGTAAAGCATCATTTTCTGCAAAATGAGGAAATTTATATGTCAAGAGGCCAAGATGTTTGACCCTAA
GGTCATTGTCCCTCTTTGTCTTTTTAAAGTCAATTATCAGTCAGACTTGATCTGTTCACTTCCTCACAATCCTCTTGTAATCTAAT
TTTGCTAATTTGAGGATAACCCTGAGCAGGTATAAAACACCAAGGCAGCCATAAACTACCTGTGTGTGGTTGGTTTACACCTTTCT
AACACTACTGGGATATTCCATCAGCTGAAACTCAAGCAAACTAGACAACAGAAG

>Oreochromis_niloticus_SWS1_3'UTR
TGCTCACACCAAAGTGACACTGAGGACTTTTAAAAGGGCGTGACTGAGCTGAAGAGGACGAATGTTTCCTGTATACTCATCGTTAC
ATGTAAATACAGTTTATACGTACACACGATCAGCTGATATCTATGTTCCTGTAGAACATACGTGTTTAGATAAAAGGCAAAAATAA
ACATTAAAAACAATGAAGTGTCTTTCTTTAGACTTTCTGAATAAA

>Metriaclima_zebra_SWS1_3'UTR
TGCTCACACCAAAGTGACACAAAGAACTTTTAAGAGGGCGTGACTGAGCTGAAGAGGACCTTCACAGGTCCGAGTTTGGAGGACGA
ATGTTTCCTGTATTCTCATCGTTATGTAAATAAAGTTTATACGTACACACGATCAGCTGATATCTATGTTCCTGTAGAACATACGT
GTTTAGATAAAAGGACAAAAATAAACATTAAAAACAGTGAAGTGTCTTTCTTTAGACTTTCTGAATAAA

>Oreochromis_niloticus_LWS_3'UTR
ATATTCACGTTCTCCATTTTGGGAAAAAAAAGAGAAATATTTTGATTTGTACAGTAAATATTTATTGTTTCCCTTTTCTTTTCTTT
TTTTTTAGTAAATATAGCTTTCTCTGGCAAATGAAAAAGCAAAAAAATACACATAAAGCAAATGAATCATTATATTAATTGTCTTT
TTTCAAATGTAAATAAA

>Metriaclima_zebra_LWS_3'UTR
ATATCCATGTTCTCCATTTTGGGAAAAAAAAGATAAATGTTTTGATTTGTACAGTAAATATTTATTGTTTCCTTTTTTCTTTCTTT
TTTTTTAGTAAATATAGCTTTCTCTGGCAAATAAAAAAGAAAAAAAATACACATAAAGCAAATGAATCATTATATTACTTGTCTTT
TTTCAAATGTAAATAAA

>Oreochromis_niloticus_RH2B_3'UTR
AAATAATACCTACATATCTCCAGCCATACTCCTTCCTGACTTTTCCCTCATTCTGAATGACTGTCTTGACATGGACATGAGAAAAT
GGATACAAGCTAATTTTCATCATTCGTTATTATTGTATGTGAAAACTTGACTCTAAATTTATTTTTGAAATGTTTGGATGGTGGTG
GGATAAAAATCCCCTTTGTATATTTGTGACATAAAGAGGTTTTCGAATATGTTAGCCTCTGTCATGATATTCTGTAAATATTTTAT
AAACAAAAAGCCTAGGAATAATATATTAAATGCATGGATGAAAAGATTTATT

>Metriaclima_zebra_RH2B_3'UTR
NNNAAATACTACCTACATATCTCCAGCCATACTCCTTCCTGACTTTTCTCTCATTCCGAATGACTGTCTTGACATGGGCATGAGAA
AATGGATACAAGTTCTATTTTTATCTGCAGGTATTTTACTGTAATTGTTCCTTACTATTGTATGTGAAAACTTGACTCTTTAAGTT
TTTTTTTTTTTTGTATGGGATAAGAATCCCCATTGTATATTTGTGACATAAAGAGGTTTTCGTAGCTTCTGTCATGATATTCTGTT
TATATTTTATAAACAAAAGCCTAGGAATAATATATTAAATGCATGCATGAAAAGATTTATT

Continued

Supplementary File 5-1 (continued)

>Oreochromis_niloticus_RH2Aalpha_3'UTR
TCATGATGGCATCTTCAGATCTATGGACATTGATGATTGCTTCCAGATTTTAAAACATCCCATTAGAGAAAGTTTTTGCCAAATGG
ATTTAAGAAGCGGACTCTGGACCAAAAACACAAAGGAATTCTATGTTATCTGTGAATTAACAGAAAGTCAAATGCATTTTCTGTTG
GGGTGATCTTGTACAGCAATTGTGTGTCTGTGTGCATGTGAGACCATTTGAGAGAGTGAATATGAGTGAACTTGTAAACTAAAAAC
ACCTCAAACCGTAAACCTGCTGAAATGTGACACTTGTACCTGATGAATGTACAATCACATGTTAATTGTTTTCATGCAATAACTTG
GGGGAGCATATATTTCACTGGTCTACCAGTGGTGTTATTTTTATGCATTATATGAAAAAGAAAAAAAAATATAGGACTGAGTAGAT
AAAATAAA

>Metriaclima_zebra_RH2Aalpha_3'UTR
NNTCCTGATGGCATCTTCAGATCTATGGACATTGATGATTGCTTCCAGATTTTAGAACATCCCATTAGAATCATTTTGCCAAATGG
ATTTAACAAATGGAGTCTGGACCAAAAACACTAAGGAATTCTATGTTATCTGTGAACTAACAGAAAGGCAAATGCATTTTCTGTTG
GGATGATGTTGTACAGCAATTGTGTGTCTGTGTGCATGTGAGCGCATTTGAGAGAGTGAATATGAGTGAACTTGTAAATAGAAAAC
ACCTCAAACTGTAAACCTGCTGAAATGTGACACTTGTACCTGATGAATGTACAATCACATGTTGATTGTTTTCATGCAATTACTTG
GGGGAGCATATATTTCACTGGTCTACCAGTGGTGTTTATTTTTATGCATTATATGAAAAGGAAAAAAAAAATAATAGGACTGAGTA
AATATAATAAAT

>Oreochromis_niloticus_RH2Abeta_3'UTR
NNNNNNNNNNNTGGCATCTTCAGATCTATGGACATTGATGATTGCTTCCAGATTTTAAAACATCCCATTAGAGAAAGTTTTTGCCA
AATGGATTTAAGAAGCGGACTCTGGACCAAAAACACGAAGGAATTCTATGTTATATGTGAACTAACAAGAAATGCATTTTCTGCTG
AGGTGATCTTGTACAGCAATTGTGCGTATGTGTGCATGTGAGCGCATTTGAGTGAGTGAGTCAATGTGTAGGAACTCAGATCCTCA
AAATGTCAAAAGAAAAACTGTAAAACTGCTGAAATGTGACACTTTTACCTGATGAATGTGCAATCACATGCTGATTGTTTTCATGT
AATTTTATGGTTGGGGATGTGTGTGTGGGGGGAGGGGGGCATACTGTATACTTAACTGGTCTACCAGTGGTGTTATTTTTATGTAT
AAAAAAAGAAAATATAGGTTTGAGTAGAAAATAAA

>Metriaclima_zebra_RH2Abeta_3'UTR
TCCTGATGGCATCTTCAGATATAAGGACACTGATGATCGCTCGCAAATTTTCAAAATTCCCATTAGAGAAATCTTTTGCCAAATGG
ATTTAAGAAGCGGACTTTGGACAAAAACAACTAAGGAATTCTATCTTATCTTATCTCTGAACTAAACAAAAAAATAGCATTTTCTG
TTGAGGTGATCTTGTACAGCAGTTGTGCGTCTGTGTGCATGTGAGAGCATTTGAGTGAGTGAGTCGATGTGTAGGAACTCAGATCC
TCAAAATGTCAAAAGAAAAACTAGTAAAACTGCTGAAATGTGACACTTTTACCTGATGAATGTGCAATCACATGCTGATTGTTTTC
ATGTAATTTTATGGTTGGGGATGTGTGTGGGGGGCATACTGTATACTTAACTGGTCTACCAGTGGTGTTATTTTTATGTATAAAAA
AATAAAATATAGACTTGAGTAGAAAATAAA

>Oreochromis_niloticus_SWS2B_3'UTR
CAATTTGGACAGCTCGCATACTGAACATAGCACAGAAATGCTAATAATATGACAAATGTATTTTTGCTATTCTTACAGATACATAG
TTTTATGCAGTAAAAATGTTTAAAAAATGCAATAAA

>Metriaclima_zebra_SWS2B_3'UTR
GTTACTCGTGCCCAATTTGGACAGCTCGCATACTGAACATAGCACAGAAATGCTAATAATGTGACAAACGTATTTTTGCTATTGTT
ACAGATACATAGTTTTATGCAGTAAAAATGTTTAACAAAAAGGAAATAAA

>Oreochromis_niloticus_SWS2A_3'UTR
GCAACACGTCAATGCTTGCTGATTTCTGAACTGTAAATTACAAAGAATGTACATGTATTTTATCATATGTAAAAATGTCATGATTT
AAGTAGGAATAATCTGCAAAATGAAAATACAAATAAA

>Metriaclima_zebra_SWS2A_3'UTR
GCAACACGTCAATGCTCGCTGATTTCTGAACTGTAAATTACAAAGAATGTACATGTATTTTATCATATGTAAAAATGTCATGATTT
AAGTAGGAATAATCTGCAAAATGAAAATGCAAATAAA

>Oreochromis_niloticus_ACTG1_PROMOTER
TAATTTGGTGGCTTCTTCTAATCTAATATTTTAGTGTTTCATATACCTTCTTTCGTTCTCGTTTTTTTTTACTGTAGCACAGGCAT
ACCACACATACACGAGCCCCGCCCCCTCAATGCTGTGTATGAAACTCTGCTGGAATGTGACCATTAAACGACTGGAAACATGGTGA
AGATTGAAACAGGTAAAGTATGCATACTGAGCATACATGGTGCTCACATACAGAATAAGAGAGTGGCTGAATTGGAGACTTAATGG
GAGTTTAAAGGGAGAGGGTGGGGTGAGAAGGGGATGAGTGAGAAGTGGACAGAAGCTACCTGCCAGCTTCTTCCATATACGGTCTT
AGTCAGAGCTGGCCTTACCCAGGACTGTGAGTTTCTCTGACTGTGTGAGTGAGCAATGCAGGGAGGGAGTGCCTACATTTGTGTGT
TGTAATGCGTGCATGAGTCATGCACGCTGCCTAGTTTTGAAATTGCTCAAGGAAAGGCGTGAGGTTGAGAAACTTCTGTCTGGGTA
GAGAACAAAGTTTAAAAGATTCTTTTAAAAGATCAATTTAATCATTTAAAACTGTGATTGAAACAAATCGGATAAAGAATACAGAA
GAAAATGTGAGCGGTGACAGACAAACACATGCTTACATACATGGCCTTAAAAAGGGAGATTCATTCATTTCCTCAGAGTGGTCACA
TGACCGGAGCAGCGCATTCTTCCCCCCATGTGACTCTCTCCAATATCTCCTTATATGGACGTCATCCAAAGGAACTCTTAAAGGGA
TGGTGGTGCCCTGGGTCTATACTTATGGCCATGGCCAGACCGAGTGAGAAACTGCCTTTGTTTCTCAGCTCTGGATTTGGACTCAA
CCTTGCGCCACTGCGACATCCAGCGTCCGTTCTCTATAACGCCGAGACTTTATCCAGGCTGTCTGCCAGATGTAACCACACTGCTG
CTGAACATTAGGACTAAAAATCCTTTCCTCTCTCGTTTCTTCTTCATACAGAAA

Continued

Supplementary File 5-1 (continued)

>Oreochromis_niloticus_AMPD3_PROMOTER
CATTTAGTCATTATCAACTCTGCACAATTACATTGAGATTTGTGAGCATTTATGAGAGGATATTGCTTTGTTTATGTTCACAAACT
GAAAAAAACAACTGCCTCATGAGCTGCTATATGGTGTGAAATCAGTTGCATGCTAAATCTGAGTGAGCATCTAGCAAACAAGTTAA
CTTCCATATGTAACTAACATTGTTAGTTTTTTTTCATGGCACCATGGGAAAAAATACTATAGCTTTGAATTTGTTCTGAGTCAGTG
TATCTCTGTAAAAGTGTTTGAATTGGGGAACAAAATGAAACAAACGCATGACAATGACTCATTACATATGAATGACCACACAGAAT
TATGTTTATTTCATGATTAATGATGCTATTGTGCTATAAAAACCGTTATAATTACTGTAAACTGGAGCTTTTTGTTCAAAGGCTGG
AAAGGTTCACTGGGTACTGTACATAATTTCAGTATGGCTCAAGTTCAGTATGGATACATCCAATTCAGTTACAATTCTTTACACAT
CTCACACACACATACAATTCAGCAAAGAGACACTAGAATAGAACAGAATAGAATAGAATAGAATAGAATAGAATATGTCCA
GCAGCAAAGTGACAGGCAGATGGAGCACGCAGCAAAAAATACAGTACAGTTAGTAAGTTATTAAAAATTAAAAAAAATATTAAAAA
CAGAAAAAGGAAAAACTGTGCAAATAACAGCAAAGATATAAACAACTATAATGTAACTATAATATAAACAACTTATTGCGTTTGTT
GGGAGCAGCGATAGTTATAAAGTCTACCAGCTGCTGCGAGGAAGGACCTGCAATAATGAATGCCCTGTCCCCAAACACGTGCACAG
GTAGATCTCCCTGTTTACAAAACACCACACAGTATCAGAGCATTTAACGTCCTGACAAGTACGGCAGACCTCCTTTGCACTCTACA
GTCAGAGACTTCAGACGCCGGCCATTTTACTGTGGCTTCTGGCCTCTTGCAGAC

>Oreochromis_niloticus_DHCR7_PROMOTER
ACGGTACTCCGTGAATAAATTGTGAGTGTGCTTTTTAAAATCCACCGTGCACGGAACAAAGGACGCAGTTATCTCTGAACGGTCAC
GGTAGATTAAGTTTGCATTATGGGAAGTGGATCTGACATATGTACTGAACAATCATACTATCAACACAGAATCCGGAAATTAACGG
TTTTGTTTACGTTATACAACGATGGGTTTCCCCCTCCCCCGTCAGATATAGCTTTCTACACGAAGCTGCTAAAACAGCGTGTTTTT
ACTTGTTTAGACTATACATATATATATATATATGAAAGCCATAAGTGTTTTAAGCTGAAGTTTTCCAGTTTTATTGTTATGGTAGC
TTTAGTGCTGCAGAACGGTGTGCATTTGTGTTACTCATATTTGTGGGGCGGTTCACAGGGTCACATTCTGGGGATCCGCCCCCTGT
TTGTGCAGATAATGCAAATTCCTGGAAAGTAAGTTTTAGTACTTGCCTTTTTTAGTTTAGGATGGGTTACAGTTAAGTTTAGAGTA
AAGCAACAGTGCATGGGAAGGAGACGTGACTGTGTCTTTGTGTGAAAGAGAGAAGCGAAAATGAAACAGATACAAAGAAGAATAAA
TGATTTCATGCATAAGGTGCCATCAGGTTTCCTGTTAAACGAGTTGGGTCTGTCTGTTATTGCACATCTCCAAGGACTTGCACTTT
ATTCCTCCAATTAGCACATTATCTGTTTATAAACGTGCGGCCTGTGGAGTAAGCTGCATGATTAGATTTGCAGTGACATTGTGACT
CTGGAGGGTTTGTTCAGCAAGTACACACAAAAGACTAGAGTCAATATGTGAACTAGAAGGTAACATCTGTTTGGTTAATAACAGGC
TGTTACTTACTGCTGGCACATTCAGTGAAAGTTACTGAGTATCTGACCGACTTACTGCAGTGGTGCACAAAGTGCAGCACATTTGT
AGATAACAGATTATAAACTGATAACAAAATTCTTCACCACAGTTGAGTGGAGCA

>Oreochromis_niloticus_ENSGAC000000020282_PROMOTER
AAATGTTTATTCAGGCCATTAAGATACTGCTGCTGTACCGTACATCCGTCCCCTGATGCCGTTTCCATTTCTTATGTAAAATTTAA
AACTAAATGAAACTAAGCTAAACGGCAGATTGGCTGATCTCAGTGGTAGCCCGCATCTGCTGATTGCTCAGCACATACCCATCTAA
GCGTCAATCTGGTTAATTTCATATAACACATGGTATTAAAGCTGCTTCAGCCAGCCCACGAGTTGCTGTGCATATGCATGCTGCTT
TGCAACCCCACTCCACCCTGTCATGCTCTGTCTGACTTAATCAGAGAGACTATGTTGGCAAATATAAGTTACTGCATGTGGAATGG
ATTCTTTTAGGTAAAGTTGTGATGACTGAATTGCATTCAGGCATTCAGCTGTATATTTCACACTAGACCTGAGAGGTGTGTTGGAC
TCTCTAGACACAAGAGCTATTCCTCCTTTAAATCATCCTATAATATATGTACCAGTAGGGCTACCTAAGAACAACTGCTCCTTTCT
TGTTATTCTCGTTGCCAGTGGATATTATTTTTGGACTCCCATCTAACTCGACAGTACGGGGTGGACTGAGTCAGTTTTTCCCTCGA
TACTACAGGTGCAAAGGGTAACATTTTCACTGCGGTTTAGCACAGAATGACCACATTATATGTGTGGGAGTCGTACAGGGTAGTTC
AATACCAGTGACTCAACGATTAGTAGGTGCCAATATTTCTGGCTTTCTAAACTCGGACCAACGAAACTGATGGATCCCACATCATT
TGCATACAAAAGAGCATGACTGCGAGCACTTTTGTGGTGCCATATCGGGGTTATTTTGAGAATATAACTGAATAATGAACATCTGA
CATGTGGCTTTCTCATCCACCAGGCAAATTTCTTTTTTTCTCCTGTCTTTCAGAGGTTCATGCGAACCTTCACATGTCACGTGCAC
TGATCTCTAGCATTAGAAGACTCAAAGCATCACCACACCATCATGTTTCTACAC

>Oreochromis_niloticus_IGFALS_PROMOTER
CTTCACTAATAGCAGATNTATAAGAATAAGAAAAGAGATCACTTATTTAACCTTTCTTTTTCCAAGTAAAAAGTTCTTTCGCTGTT
ACAGCAACAGAGNCACAGAAATGAAAGTGCTCCTTTTGGCTGTTTTCCACTGTGTTATGCGCAACAGATGAAGGAGGGTCCAACAT
TAGTCATGTCCTGAGGCCCACTACAGCAGCACAGGAGGAGTAATCTTTGGTTTTACAGCTTGTTTTTTTCTTCTTCTTTTGATAGA
GCTATTTCTTCACCGAAGTTACTACATTCAGCTGGTTTTTATGTACTTCTCTTAGACTCATTTTAGAATCCTGGTGGTGTTTGGCT
GTCATTGCTTTTCTAAAATGAAGCAGTTAATACCGCACACAGTCTTTTTAGTGTTGGTTTGTCACATAGCGTTTTATGTTTGAAAA
GACAACAAGATAAGTTGTGTGGTTTCAAAGGTGAATCAGCTTGGACTGTGAATTGAATACCAGCCAATTAAACAAAAAAAACACAG
AGAAAAAACAAAGTACTAAAGTCTGAGTCATAATCACTTTAACAAGCACAGCAACAAATTTTCATCTTCACTTGCTAAATGGGCAT
GTTTAATTGTTTTTACTGTGATGCTGAAAAGGTGTAGGTCATTGGATGGGATGAGAGTGACTAGATATATGTGTTTTACTGTTAAT
GATGCACAATAGATAGATAGATGGACAGGCAATGTAGTATAGTATATATTTCCAAATCAGTTAAACCTAATAAAAATCCCTGAAAA
CCCAGTTCAGTTTAAACTTGTATGATTTCTGTATCCAGATAATGATTACAAAAGTCTAAGTTAAACAGGCCTAACATAGCTGCAGC
AGGATAAGCCTGTGTGTTCATTAGCTACAGTGTATGTGTTCTAATACTTTATCTCTCTAACTGGCAATAACTCTGATGAGCATCCA
GCTGCTTAATTTAGATTAACATCTTGCTTTCTAGGCAATTTCTCTGCTGTTAAA

Continued

Supplementary File 5-1 (continued)

>Oreochromis_niloticus_KCNJ9_PROMOTER
TTCTGGACCTCAGTTCAACCAGACTGAAACGCATCACTATTTATCTCTTTGTTTCTGATAAAAATCTCCAGAGATTTAGAGATATT
GACTTACTCCGGAGTCTGGATGAACATTTTCTCCATTCAGCCCTGCACCCTCTTCTCCTCCCCATCTGTCTCGCAGTAGTAGTAGT
GAATCTTTTCTAATTTAGACCATACTGCAGAGCTTCACCCCTCAGCAATTTTGACAGTCTTGCTCATTAAGCTCTGCTCTGTTTGC
TCCAATCTTTGCCGTACATCTTCATACAATAAAAAGAAAGAAAGCCTAACTAGGGAACTAGGGAAATGTGCAGTTTATATGACCCT
TTGGTGTTTTAACTCCAACTTGTGATCAAGAATTGATGTCATCATTGCAAAGGACAAAAGTCCTACAAAGCAGAGCCAGACCTTTT
TCGCAGAAAAGGCAGATTTTGTTTTACCTTAAAATAAGTCTCCATAGCTTGAGTAGCATGCACTATGATTGATATTCATTGATTAA
AGCGTATGTGATAGCGGTTTAGTGTCATAATTTCAGTTCTGGCTGGGTGCATCCACAGTAAAAAGCCTATAAAACAAAACAAAACA
CAGATGAAGTCCAGTGTTATAGCTTTTCTTCTCCAAGCGTTTGATGATGTTGTCCCGTCGCACTTCATCTGCTCCGTTTTATTCAC
CTGTCATTTTTCATTCATGTTTTCCCTTCTTCACCAGTCTGACTTTCCTTTGAGCCACGCTGCATCTCCCTCAGTAGTTTTGTCCC
TTGCTACCTCCCTGTGTTTATGCTCACGTCTCTCGAAGTGCCTTCATCCAATTCTCATTGTCTGACTGTAGTCAGAGTTGTAGCTT
TAAAACCAAAACATTTCAGAAAGGGAGCCTCATGCATGTTTTTGAAAGTGTTTAATGGAACAATAAAGAGCTGTCTTCAACACTGT
AAAGGATCAAAAGGCGATGACACTTTTCCTCCCATTTGTTAGCCTCACATATTT

>Oreochromis_niloticus_REEP1_PROMOTER
GGGATAAGAATTTCGAGGATAGGCAATATCCCTGTCCTTGAGATCACTTTTTAGTTTTTATGAGGTCATTTTAATAAGGTTTGACC
TTCAGTTGTAAATGTCAATTGTGAGCAAGATTACAGTCAATACAATCTGAAATATTTGGACTCTTGCTCATTGTTTTAGGCCACAC
TAGCTACCGTGCATCTTCCTCAAATAACCCACTGTCACTTGATTTGTGCTCATTTGATTAATTGTAACATTTTCCTTTAAAAATGT
TACACCATAGAATAGTAGTAACATCAGGACCCTGCATCCTGAAGTGGGCCTGTGTGTATGTAAAGAGCATGACGTGTTAATGGGTA
ACAGGTGAAAACACCATGTTCTCTTTACAGTTGTTTAATAAAGTTCTTTAATCTTTTCTTTCAAGAGTGTTCAGACTAAAATATTT
GTTCCTGTCTGCAGCTCTGAGGCAGTGTGATAGATAATACTCTCTACTATTCGGACATAAACAAAACAAGCAATCAAGAAATTTAC
TTGTAGATAAATCGCTAATTAAAATGTTTTCGTTGCAGTAATGCGTGTGTGTGTAGCTTCATTTAAATGACCACAAAATTATTCAT
TCAAACACATATCGGCCATGTACAATTACCGTATTATTCCAAGTAGGAGGTAGTTTCTGGATCTTTGAGCTTTACGGGAATTTGGT
TGCAGTATTAAAAGGTAAATTTGAAAGGGATTGCAGAATACAAACGCTCAAAAGGAGCACGGAGAGGGGGAACAGGAGATATTTGT
CTCCTTTGCAACCCTGCATAAAATCGACACATCCACACACCCTTGTGGGCTGTCTGCTGTCTGTGCAATGCAACACTGCCATCTAT
GGATAAGAGCTGTAGGCCGCTCACTGTTTATCCCGCATCACGTTGCCGCCTGTGTGCGTCCACACGCAGCCGCCTGGACGCGCCTG
TCATCTGGAGGAAACTCCTCCCTGCAGACAGCGCGTCCTGTCGCCTTCCCAGCG

Bibliography

Albertson RC, Markert JA, Danley PD, and Kocher TD. 1999. Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. *Proc Natl Acad Sci U S A* 96:5107-10.

Albertson RC, Streelman JT, and Kocher TD. 2003. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proc Natl Acad Sci U S A* 100:5252-5257.

Allender CJ, Seehausen O, Knight ME, Turner GF, and Maclean N. 2003. Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. *Proc Natl Acad Sci U S A* 100:14074-9.

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403-10.

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in Drosophila. *Nature* 437:1149-1152.

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

Arendt J and Reznick D. 2008. Convergence and parallelism: what have we learned about the genetics of adaptation? *Trends Ecol Evol* 23:26-32.

Arora A, McKay GJ, and Simpson DA. 2007. Prediction and verification of miRNA expression in human and rat retinas. *Invest Ophthalmol Vis Sci* 48:3962-7.

Asenjo AB, Rim J, and Oprian DD. 1994. Molecular determinants of human red/green color discrimination. *Neuron* 12:1131-8.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, and Johnson EA. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* 3:e3376.

Baker DL, Dave V, Reed T, and Periasamy M. 1996. Multiple Sp1 binding sites in the cardiac/slow twitch muscle sarcoplasmic reticulum Ca2+-ATPase gene promoter are required for expression in Sol8 muscle cells. *J Biol Chem* 271:5921-8.

Balmer JE and Blomhoff R. 2002. Gene expression regulation by retinoic acid. *J Lipid Res* 43:1773-1808.

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281-97.

Becker RA, Wilks AR, Brownrigg R, and Minka TP. 2010. maps: Draw geographical maps.

Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, Levine M, Rubin GM, and Eisen MB. 2002. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. *Proc Natl Acad Sci U S A* 99:757-762.

Birney EStamatoyannopoulos JADutta AGuigo RGingeras TRMargulies EHWeng ZSnyder MDermitzakis ETThurman RE, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799-816.

Boettner EA and Wolter JR. 1962. Transmission of the Ocular Media. *Investigative Ophthalmology & Visual Science* 1:776-783.

Boffelli D, McAuliffe J, Ovcharenko D, Lewis KD, Ovcharenko I, Pachter L, and Rubin EM. 2003. Phylogenetic Shadowing of Primate Sequences to Find Functional Regions of the Human Genome. *Science* 299:1391-1394.

Bowmaker JK. 1995. The visual pigments of fish. *Prog Retin Eye Res* 15:1-31.

Bowmaker JK, Govardovskii VI, Shukolyukov SA, Zueva LV, Hunt DM, Sideleva VG, and Smirnova OG. 1994. Visual pigments and the photic environment: the cottoid fish of Lake Baikal. *Vision Res* 34:591-605.

Bowmaker JK, Heath L, Wilkie S, and Hunt DM. 1997. Visual pigments and oil droplets from six classes of photoreceptor in the retinas of birds. *Vision Res* 37.

Brichard P. 1978. *Fishes of Lake Tanganyika*. Neptune City, NJ: TFH Publications Inc.

Brichard P. 1989. *Cichlids of Lake Tanganyika*. Neptune City: TFH Publications, Inc.

Britten RJ and Davidson EH. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev Biol* 46:111-38.

Brock G, Pihur V, Datta S, and Datta S. 2008. clValid: an R package for cluster validation. *J Stat Softw* 25:1-22.

Browman H and Hawryshyn C. 1994a. Retinoic Acid Modulates Retinal Development in the Juveniles of a Teleost Fish. *J Exp Biol* 193:191-207.

Browman HI and Hawryshyn CW. 1994b. The developmental trajectory of ultraviolet photosensitivity in rainbow trout is altered by thyroxine. *Vision Res* 34:1397-406.

Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, and Batzoglou S. 2003. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* 13:721-31.

Calissano M, Diss JK, and Latchman DS. 2007. Post-transcriptional regulation of the Brn-3b transcription factor in differentiating neuroblastoma cells. *FEBS Lett* 581:2490-6.

Cammas L, Trensz F, Jellali A, Ghyselinck NB, Roux MJ, and Dollé P. 2010. Retinoic Acid Receptor (RAR)-α Is Not Critically Required for Mediating Retinoic Acid Effects in the Developing Mouse Retina. *Investigative Ophthalmology & Visual Science* 51:3281-3290.

Carleton KL. 2009. Cichlid fish visual systems: mechanisms of spectral tuning. *Integrative Zoology* 4:75-86.

Carleton KL, Harosi FI, and Kocher TD. 2000. Visual pigments of African cichlid fishes: evidence for ultraviolet vision from microspectrophotometry and DNA sequences. *Vision Res* 40:879-90.

Carleton KL, Hofmann CM, Klisz C, Patel Z, Chircus LM, Simenauer LH, Soodoo N, Albertson RC, and Ser JR. 2010. Genetic basis of differential opsin gene expression in cichlid fishes. *J Evol Biol* 23:840-53.

Carleton KL and Kocher TD. 2001. Cone opsin genes of african cichlid fishes: tuning spectral sensitivity by differential gene expression. *Mol Biol Evol* 18:1540-50.

Carleton KL, Parry JW, Bowmaker JK, Hunt DM, and Seehausen O. 2005a. Colour vision and speciation in Lake Victoria cichlids of the genus Pundamilia. *Mol Ecol* 14:4341-53.

Carleton KL, Spady TC, and Cote RH. 2005b. Rod and cone opsin families differ in spectral tuning domains but not signal transducing domains as judged by saturated evolutionary trace analysis. *J Mol Evol* 61:75-89.

Carleton KL, Spady TC, and Kocher TD. 2006. Visual communication in East African cichlid fishes: diversity in a phylogenetic context. In *Communication in Fishes*, edited by Ladich F, Collin SP, Moller P and Kapoor BG. Enfield: Science Publishers.

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

Carroll SB. 2005. Evolution at two levels: on genes and form. *PLoS Biol* 3:e245.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134:25-36.

Chan YF, Marks ME, Jones FC, Villarreal G, Jr., Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327:302-5.

Chen K and Rajewsky N. 2006. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* 38:1452-1456.

Chen K and Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 8:93-103.

Chen Y and Stephan W. 2006. Weak selection on noncoding gene features. In *Evolutionary Genetics: Concepts and Case Studies*. Oxford: Oxford University Press.

Chessel D, Dufor AB, and Thioulouse J. 2004. The ade4 package - I: One-table methods. *R News* 4:5-10.

Chiba H, Clifford J, Metzger D, and Chambon P. 1997. Specific and Redundant Functions of Retinoid X Receptor/Retinoic Acid Receptor Heterodimers in Differentiation, Proliferation, and Apoptosis of F9 Embryonal Carcinoma Cells. *The Journal of Cell Biology* 139:735-747.

Chinen A, Hamaoka T, Yamada Y, and Kawamura S. 2003. Gene duplication and spectral diversification of cone visual pigments of zebrafish. *Genetics* 163:663-75.

Coddington JA. 1988. Cladistic tests of adaptational hypotheses. *Cladistics* 4:3-22.

Conte I, Carrella S, Avellino R, Karali M, Marco-Ferreres R, Bovolenta P, and Banfi S. 2010. miR-204 is required for lens and retinal development via Meis2 targeting. *Proceedings of the National Academy of Sciences* 107:15491-15496.

Couronne O, Poliakov A, Bray N, Ishkhanov T, Ryaboy D, Rubin E, Pachter L, and Dubchak I. 2003. Strategies and Tools for Whole-Genome Alignments. *Genome Research* 13:73-80.

Cowing JA, Poopalasundaram S, Wilkie SE, Bowmaker JK, and Hunt DM. 2002. Spectral Tuning and Evolution of Short Wave-Sensitive Cone Pigments in Cottoid Fish from Lake Baikal,Ä†. *Biochemistry* 41:6019-6025.

Coyne JA and Orr HA. 2004. *Speciation*. Sunderland, MA: Sinauer Associates.

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

Darwin C. 1859. *On the origin of species*. London, UK: Murray, J.

Decembrini S, Bressan D, Vignali R, Pitto L, Mariotti S, Rainaldi G, Wang X, Evangelista M, Barsacchi G, and Cremisi F. 2009. MicroRNAs couple cell fate and developmental timing in retina. *Proc Natl Acad Sci U S A* 106:21179-84.

Di Palma F, Kidd C, Borowsky R, and Kocher TD. 2007. Construction of bacterial artificial chromosome libraries for the Lake Malawi cichlid (Metriaclima zebra), and the blind cavefish (Astyanax mexicanus). *Zebrafish* 4:41-7.

Díaz-Uriarte R and Garland Jr. T. 2007. PHYLOGR: Functions for phylogenetically based statistical analyses. R package version 1.0.6.

Dietrich MR. 2006. From Mendel to molecules: a brief history of evolutionary genetics. In *Evolutionary Genetics: Concepts and Case Studies*, edited by Fox CW and Wolf JB. New York,: Oxford University Press.

Dobberfuhl AP, Ullmann JF, and Shumway CA. 2005. Visual acuity, environmental complexity, and social organization in African cichlid fishes. *Behav Neurosci* 119:1648-55.

Dobzhansky T. 1937. *Genetics and the origin of species*. New York: Columbia University Press.

Dowling JE. 1987. *The Retina: An Approachable Part of the Brain*: Belknap Press of Harvard University Press.

Dräger UC and McCaffery P. 1997. Retinoic acid and development of the retina. *Progress in Retinal and Eye Research* 16:323-351.

Duftner N, Koblmuller S, and Sturmbauer C. 2005. Evolutionary relationships of the Limnochromini, a tribe of benthic deepwater cichlid fish endemic to Lake Tanganyika, East Africa. *J Mol Evol* 60:277-89.

Durston AJ, Timmermans JPM, Hage WJ, Hendriks HFJ, de Vries NJ, Heideveld M, and Nieuwkoop PD. 1989. Retinoic acid causes an anteroposterior transformation in the developing central nervous system. *Nature* 340:140-144.

Ebert BL, Firth JD, and Ratcliffe PJ. 1995. Hypoxia and Mitochondrial Inhibitors Regulate Expression of Glucose Transporter-1 via Distinct Cis-acting Sequences. *Journal of Biological Chemistry* 270:29083-29089.

Ellegren H and Sheldon BC. 2008. Genetic basis of fitness differences in natural populations. *Nature* 452:169-175.

Elmer KR, Reggio C, Wirth T, Verheyen E, Salzburger W, and Meyer A. 2009. Pleistocene desiccation in East Africa bottlenecked but did not extirpate the adaptive radiation of Lake Victoria haplochromine cichlid fishes. *Proceedings of the National Academy of Sciences* 106:13404-13409.

Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, and Jarvela I. 2002. Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 30:233-237.

Evans RM. 1988. The steroid and thyroid hormone receptor superfamily. *Science* 240:889-895.

Ewing B, Hillier L, Wendl MC, and Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175-85.

Felsenstein J. 1985. Phylogenies and the Comparative Method. *The American Naturalist* 125:1.

Fernald RD. 1981. Chromatic organization of a cichlid fish retina. *Vision Res* 21:1749-53.

Fernald RD and Hirata NR. 1977. Field study of *Haplochromis burtoni*: habitats and co-habitant. *Environmental Biology of Fishes* 2:299-308.

Fernald RD and Liebman PA. 1980. Visual receptor pigments in the African cichlid fish, Haplochromis burtoni. *Vision Res* 20:857-64.

Fisher RA. 1930. *The genetical theory of natural selection*. Oxford: Oxford University Press.

Fox CW and Wolf JB. 2006. *Evolutionary Genetics: Concepts and Case Studies*. New York: Oxford University Press.

Fox J. *car: companion to applied regression. R package version 1.2-9* 2008. Available from http://CRAN.R-project.org/web/packages/car/.

Fryer G. 1959. Some aspects of evolution in Lake Nyasa. *Evolution* 13:440-451.

Fryer G and Iles TD. 1972. *The cichlid fishes of the Great Lakes of Africa: Their biology and evolution*. Neptune City: T. F. H. Publications.

Fuentes C and Casella G. 2009. Testing for the existence of clusters. *SORT* 33:115-146.

Fujieda H, Bremner R, Mears AJ, and Sasaki H. 2009. Retinoic acid receptor-related orphan receptor α regulates a subset of cone genes during mouse retinal development. *Journal of Neurochemistry* 108:91-101.

Fujimura K and Okada N. 2007. Development of the embryo, larva, and early juvenile of Nile tilapia *Oreochromis niloticus* (Pisces: Cichlidae): Developmental staging system. *Devel Growth Differ* 49:301-324.

Fujimura K and Okada N. 2008. Bone development in the jaw of Nile tilapia *Oreochromis niloticus* (Pisces: Cichlidae). *Devel Growth Differ* 50:339-355.

Fuller RC, Carleton KL, Fadool JM, Spady TC, and Travis J. 2004. Population variation in opsin expression in the bluefin killifish, *Lucania goodei*: a real-time PCR study. *J Comp Physiol A* 190:147-154.

Fuller RC, Carleton KL, Fadool JM, Spady TC, and Travis J. 2005. Genetic and environmental variation in the visual properties of bluefin killifish, Lucania goodei. *J Evol Biol* 18:516-23.

Furukawa T, Morrow EM, and Cepko CL. 1997. Crx, a Novel otx-like Homeobox Gene, Shows Photoreceptor-Specific Expression and Regulates Photoreceptor Differentiation. *Cell* 91:531-541.

Garièpy J-L, Bauer DJ, and Cairns RB. 2001. Selective breeding for differential aggression in mice provides evidence for heterochrony in social behaviours. *Animal Behaviour* 61:933-947.

Garland Jr. T and Díaz-Uriarte R. 1999. Polytomies and phylogenetically independent contrasts: an examination of the bounded degrees of freedom approach. *Syst Biol* 48:547-558.

Garland Jr. T, Dickerman AW, Janis CM, and Jones JA. 1993. Phylogenetic Analysis of Covariance by Computer Simulation. *Syst Biol* 42:265-292.

Gavrilets S. 2004. *Fitness landscapes and the origin of species*. Princeton, NJ: Princeton University Press.

Genner MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, and Turner GF. 2007. Age of cichlids: new dates for ancient lake fish radiations. *Mol Biol Evol* 24:1269-82.

GenomeWiki. *Opsin evolution: LWS PhyloSNPs* 2009 03/23/2010. Available from http://genomewiki.ucsc.edu/index.php/Opsin_evolution:_LWS_PhyloSNPs.

Giguère V, Shago M, Zirngibl R, Tate P, Rossant J, and Varmuza S. 1990. Identification of a novel isoform of the retinoic acid receptor gamma expressed in the mouse embryo. *Molecular and Cellular Biology* 10:2335-2340.

Giguère V, Tini M, Flock G, Ong E, Evans RM, and Otulakowski G. 1994. Isoform-specific amino-terminal domains dictate DNA-binding properties of ROR alpha, a novel family of orphan hormone nuclear receptors. *Genes & Development* 8:538-553.

Gilbert SF. 1997. *Developmental Biology*. 5th ed. Sunderland: Sinauer Associates, Inc.

Gompel N and Prud'homme B. 2009. The causes of repeated genetic evolution. *Developmental Biology* 332:36-47.

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, and Carroll SB. 2004. Chance caught on the wing: cis-regulatory evolution and the origin of pigmentation patterns in Drosophila. *Nature* 433:481-487.

Gopal V, Fuentes C, and Casella G. *bayesclust: Tests/Searches for significant clusters in genetic data. R package version 2.1* 2009. Available from http://cran.r-project.org/web/packages/bayesclust/.

Gould SJ. 1977. *Ontogeny and Phylogeny*. Cambridge: Harvard University Press.

Govardovskii VI, Fyhrquist N, Reuter T, Kuzmin DG, and Donner K. 2000. In search of the visual pigment template. *Vis Neurosci* 17:509-28.

Grafen A. 1989. The Phylogenetic Regression. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 326:119-157.

Graur D and Li W-H. 2000. *Fundamentals of Molecular Evolution*. Sunderland, MA: Sinauer Associates.

Green P. *Phrap: phragment assembly program* 1994. Available from http://www.phrap.org.

Greenwood PH. 1974. Cichlid fishes of Lake Victoria, East Africa: the biology and evolution of a species flock. *Bull Br Mus Nat Hist Zool Suppl* 6:1-134.

Griffiths-Jones S, Saini HK, Dongen Sv, and Enright AJ. 2007. miRBase: tools for microRNA genomics. *Nucleic Acids Research*:gkm952.

Grosenick L, Clement TS, and Fernald RD. 2007. Fish can infer social rank by observation alone. *Nature* 445:429-32.

Guerin MB, McKernan DP, O'Brien CJ, and Cotter TG. 2006. Retinal ganglion cells: dying to survive. *Int J Dev Biol* 50:665-74.

Gumucio DL, Shelton DA, Zhu W, Millinoff D, Gray T, Bock JH, Slightom JL, and Goodman M. 1996. Evolutionary strategies for the elucidation of cis and trans factors that regulate the developmental switching programs of the beta-like globin genes. *Mol Phylogenet Evol* 5:18-32.

Hackler L, Jr., Wan J, Swaroop A, Qian J, and Zack DJ. 2010. MicroRNA profile of the developing mouse retina. *Invest Ophthalmol Vis Sci* 51:1823-31.

Hagedorn M and Fernald RD. 1992. Retinal growth and cell addition during embryogenesis in the teleost, Haplochromis burtoni. *J Comp Neurol* 321:193-208.

Hagedorn M, Mack AF, Evans B, and Fernald RD. 1998. The embryogenesis of rod photoreceptors in the teleost fish retina, Haplochromis burtoni. *Brain Res Dev Brain Res* 108:217-27.

Halligan DL, Eyre-Walkder A, Andolfatto P, and Keightley PD. 2004. Patterns of evolutionary constraints in intronic and intergenic DNA of Drosophila. *Genome Res* 14:273-279.

Halstenberg S, Lindgren KM, Samagh SP, Nadal-Vicens M, Balt S, and Fernald RD. 2005. Diurnal rhythm of cone opsin expression in the teleost fish Haplochromis burtoni. *Vis Neurosci* 22:135-41.

Harmon L, Weir J, Brock C, Glor R, Challenger W, and Hunt G. *geiger: Analysis of evolutionary diversification. R package version 1.3-1* 2009. Available from http://CRAN.R-project.org/package=geiger.

Hartl DL and Clark AG. 2006. *Principles of Population Genetics*. 4th ed. Sunderland, MA: Sinaur Associates, Inc.

He L and Hannon GJ. 2004. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 5:522-531.

Hittinger CT, Stern DL, and Carroll SB. 2005. Pleiotropic functions of a conserved insect-specific Hox peptide motif. *Development* 132:5261-5270.

Hoekstra HE and Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61:995-1016.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, and Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313:101-4.

Hofmann CM and Carleton KL. 2009. Gene duplication and differential gene expression play an important role in the diversification of visual pigments in fish. *J Int Comp Biol* 49:630-643.

Hofmann CM, O'Quin KE, Marshall NJ, and Carleton KL. 2010a. The relationship between lens transmission and opsin gene expression in cichlids from Lake Malawi. *Vision Res* 50:357-363.

Hofmann CM, O'Quin KE, Marshall NJ, Cronin TW, Seehausen O, and Carleton KL. 2009. The eyes have it: regulatory and structural changes both underlie cichlid visual pigment diversity. *PLoS Biol* 7:e1000266.

Hofmann CM, O'Quin KE, Smith AR, and Carleton KL. 2010b. Plasticity of opsin gene expression in cichlids from Lake Malawi. *Molecular Ecology* 19:2064-2074.

Hsia CC and McGinnis W. 2003. Evolution of transcription factor function. *Curr Opin Genet Dev* 13:199-206.

Huber R, van Staaden MJ, Kaufman LS, and Liem KF. 1997. Microhabitat use, trophic patterns, and the evolution of brain structure in African cichlids. *Brain Behav Evol* 50:167-82.

Huelsenbeck JP and Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754-5.

Hunt DM, Dulai KS, Partridge JC, Cottrill P, and Bowmaker JK. 2001. The molecular basis for spectral tuning of rod visual pigments in deep-sea fish. *J Exp Biol* 204:3333-44.

Hunt DM, Fitzgibbon J, Slobodyanyuk SJ, and Bowmaker JK. 1996. Spectral tuning and molecular evolution of rod visual pigments in the species flock of cottoid fish in Lake Baikal. *Vision Res* 36:1217-1224.

Hutvagner G and Zamore PD. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* 297:2056-60.

Huxley J. 1942. *Evolution: The Modern Synthesis*. London: Allen & Unwin.

Hyatt GA and Dowling JE. 1997. Retinoic acid: a key molecule for eye and photoreceptor development. *Invest Ophthalmol Vis Sci* 38:1471-1475.

Jacobs GH. 1996. Primate photopigments and primate color vision. *Proceedings of the National Academy of Sciences of the United States of America* 93:577-581.

Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, and Carroll SB. 2008. The evolution of gene regulation underlies a morphological difference between two Drosophila sister species. *Cell* 132:783-93.

Jessen TH, Weber RE, Fermi G, Tame J, and Braunitzer G. 1991. Adaptation of bird hemoglobins to high altitudes: demonstration of molecular mechanism by protein engineering. *Proc Natl Acad Sci U S A* 88:6519-22.

Jia L, Oh ECT, Ng L, Srinivas M, Brooks M, Swaroop A, and Forrest D. 2009. Retinoid-related orphan nuclear receptor RORβ is an early-acting factor in rod photoreceptor development. *Proceedings of the National Academy of Sciences* 106:17534-17539.

Jones NC, Rigby PW, and Ziff EB. 1988. Trans-acting protein factors and the regulation of eukaryotic transcription: lessons from studies on DNA tumor viruses. *Genes & Development* 2:267-281.

Jordan R, Howe D, Juanes F, Stauffer JJ, and Loew E. 2004. Ultraviolet radiation enhances zooplanktivory rate in ultraviolet sensitive cichlids. *African Journal of Ecology* 42:228-231.

Jordan R, Kellogg K, Howe D, Juanes F, Stauffer Jr J, and Loew E. 2006. Photopigment spectral absorbance of Lake Malaŵi cichlids. *Journal of Fish Biology* 68:1291-1299.

Karali M, Peluso I, Marigo V, and Banfi S. 2007. Identification and characterization of microRNAs expressed in the mouse eye. *Invest Ophthalmol Vis Sci* 48:509-15.

Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM, Habegger L, Rozowsky J, Shi M, Urban AE, et al. 2010. Variation in transcription factor binding among humans. *Science* 328:232-5.

Kassam DD, Adams DC, Hori M, and Yamaoka K. 2003. Morphometric analysis on ecomorphologically equivalent cichlid species from Lakes Malawi and Tanganyika. *Journal of Zoology* 260:153-157.

Kastner P, Grondona JM, Mark M, Gansmuller A, LeMeur M, Decimo D, Vonesch J-L, DollÈ P, and Chambon P. 1994. Genetic analysis of RXR[alpha] developmental function: Convergence of RXR and RAR signaling pathways in heart and eye morphogenesis. *Cell* 78:987-1003.

Kastner P, Krust A, Mendelsohn C, Garnier JM, Zelent A, Leroy P, Staub A, and Chambon P. 1990. Murine isoforms of retinoic acid receptor gamma with specific patterns of expression. *Proceedings of the National Academy of Sciences* 87:2700-2704.

Katagiri T, Asakawa S, Minagawa S, Shimizu N, Hirono I, and Aoki T. 2001. Construction and characterization of BAC libraries for three fish species; rainbow trout, carp and tilapia. *Anim Genet* 32:200-4.

Kato M, Putta S, Wang M, Yuan H, Lanting L, Nair I, Gunn A, Nakagawa Y, Shimano H, Todorov I, et al. 2009. TGF-beta activates Akt kinase through a microRNA-dependent amplifying circuit targeting PTEN. *Nat Cell Biol* 11:881-9.

Katoh K, Misawa K, Kuma K, and Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059-66.

Katoh K and Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9:286-298.

Keightley PD and Gaffney DJ. 2003. Functional constraints and frequency of deleterious mutations in noncoding DNA of rodents. *Proc Natl Acad Sci U S A* 100:13402-13406.

Kidd MR, Kidd CE, and Kocher TD. 2006. Axes of differentiation in the bower-building cichlids of Lake Malawi. *Mol Ecol* 15:459-78.

King MC and Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107-16.

Kloc M, Bilinski S, Pui-Yee Chan A, and Etkin LD. 2000. The Targeting of Xcat2 mRNA to the Germinal Granules Depends on a cis-Acting Germinal Granule Localization Element within the 3'UTR. *Dev Biol* 217:221-229.

Kloosterman WP, Steiner FA, Berezikov E, de Bruijn E, van de Belt J, Verheul M, Cuppen E, and Plasterk RH. 2006. Cloning and expression of new microRNAs from zebrafish. *Nucleic Acids Res* 34:2558-69.

Koblmüller S, Egger B, Sturmbauer C, and Sefc KM. 2007. Evolutionary history of Lake Tanganyika's scale-eating cichlid fishes. *Mol Phylogenet Evol* 44:1295-305.

Koblmüller S, Schliewen UK, Duftner N, Sefc KM, Katongo C, and Sturmbauer C. 2008. Age and spread of the haplochromine cichlid fishes in Africa. *Mol Phylogenet Evol* 49:153-69.

Kocher TD. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet* 5:288-98.

Kocher TD, Conroy JA, McKaye KR, and Stauffer JR. 1993. Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. *Mol Phylogenet Evol* 2:158-65.

Kocher TD, Conroy JA, McKaye KR, Stauffer JR, and Lockwood SF. 1995. Evolution of NADH dehydrogenase subunit 2 in east African cichlid fish. *Mol Phylogenet Evol* 4:420-32.

Kolb H. 2003. How the Retina Works. *American Scientist* 91:28-35.

Konings A. 2007. *Malawi cichlids in their natural habitat*. El Paso, TX: Cichlid Press.

Korenbrot JI and Fernald RD. 1989. Circadian rhythm and light regulate opsin mRNA in rod photoreceptors. *Nature* 337:454-7.

Kuwamura T. 1986. Parental care and mating systems of cichlid fishes in Lake Tanganyika: a preliminary field survey. *J Ethology* 4:129-146.

Lammer EJ, Chen DT, Hoar RM, Agnish ND, Benke PJ, Braun JT, Curry CJ, Fernhoff PM, Grix AW, Lott IT, et al. 1985. Retinoic Acid Embryopathy. *New England Journal of Medicine* 313:837-841.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860-921.

Latchman DS. 2004. *Eukaryotic transcription factors*. 4th ed. Amsterdam ; Boston: Elsevier/Academic Press.

Lee BY, Lee WJ, Streelman JT, Carleton KL, Howe AE, Hulata G, Slettan A, Stern JE, Terai Y, and Kocher TD. 2005. A second-generation genetic linkage map of tilapia (Oreochromis spp.). *Genetics* 170:237-44.

Lee W, Conroy J, and Howell WH. 1995. Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution* 41:54-66.

Lemos B, Araripe LO, Fontanillas P, and Hartl DL. 2008. Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc Natl Acad Sci U S A* 105:14471-6.

Leroy P, Krust A, Zelent A, Mendelsohn C, Garnier JM, Kastner P, Dierich A, and Chambon P. 1991. Multiple isoforms of the mouse retinoic acid receptor alpha are generated by alternative splicing and differential induction by retinoic acid. *EMBO* 10:59-69.

Letovsky J and Dynan WS. 1989. Measurement of the binding of transcription factor Sp1 to a single GC box recognition sequence. *Nucleic Acids Res* 17:2639-53.

Levine JS and MacNichol Jr. EF. 1979. Visual pigments in teleost fishes: effects of habitat, microhabitat, and behavior on visual system evolution. *Sens. Processes* 3:95-131.

Levine JS, MacNichol Jr. EF, Kraft T, and Collns BA. 1979. Intraretinal distribution of cone pigments in certain teleost fishes. *Science* 204:523-526.

Levine M and Tjian R. 2003. Transcription regulation and animal diversity. *Nature* 424:147-51.

Lewis BP, Burge CB, and Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120:15-20.

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, and Burge CB. 2003. Prediction of mammalian microRNA targets. *Cell* 115:787-98.

Lewontin RC. 1991. Twenty-five years ago in genetics: electrophoresis in the development of evolutionary genetics: milestone or millstone? *Genetics* 128:657-662.

Li E, Sucov HM, Lee KF, Evans RM, and Jaenisch R. 1993. Normal development and growth of mice carrying a targeted disruption of the alpha 1 retinoic acid receptor gene. *Proc Natl Acad Sci U S A* 90:1590-1594.

Li QJ, Chau J, Ebert PJ, Sylvester G, Min H, Liu G, Braich R, Manoharan M, Soutschek J, Skare P, et al. 2007. miR-181a is an intrinsic modulator of T cell sensitivity and selection. *Cell* 129:147-61.

Librado P and Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.

Lisney TJ, Studd E, and Hawryshyn CW. 2010. Electrophysiological assessment of spectral sensitivity in adult Nile tilapia *Oreochromis niloticus*: evidence for violet sensitivity. *J Exp Biol* 213:1453-1463.

Llave C, Xie Z, Kasschau KD, and Carrington JC. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* 297:2053-6.

Loew ER, McFarland WN, Mills EL, and Hunter D. 1993. A chromatic action spectrum for planktonic predation by juvenile yellow perch, *Perca flavescens*. *Can J Zool* 71:384-386.

Loh Y-HE, Katz LS, Mims MC, Kocher TD, Yi SV, and Streelman JT. 2008. Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. *Genome Biol* 9:R113.

Loh YH, Yi SV, and Streelman JT. 2010. Evolution of microRNAs and the diversification of species. *Genome Biol Evol* doi: 10.1093/gbe/evq085.

Lohnes D, Mark M, Mendelsohn C, Dolle P, Dierich A, Gorry P, Gansmuller A, and Chambon P. 1994. Function of the retinoic acid receptors (RARs) during development (I). Craniofacial and skeletal abnormalities in RAR double mutants. *Development* 120:2723-2748.

Löhr U and Pick L. 2005. Cofactor-interaction motifs and the cooption of a homeotic hox protein into the segmentation pathway of *Drosophila melanogaster*. *Curr Biol* 15:643-649.

Losos JB. 1994. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. *Syst Biol* 43:117-123.

Lufkin T, Lohnes D, Mark M, Dierich A, Gorry P, Gaub MP, LeMeur M, and Chambon P. 1993. High postnatal lethality and testis degeneration in retinoic acid receptor alpha mutant mice. *Proc Natl Acad Sci U S A* 90:7225-7229.

Luo J, Sucov HM, Bader J-A, Evans RM, and Giguère V. 1996. Compound mutants for retinoic acid receptor (RAR)β and RARα1 reveal developmental functions for multiple RARβ isoforms. *Mech Devlop* 55:33-44.

Lynch M and Conery JS. 2000. The Evolutionary Fate and Consequences of Duplicate Genes. *Science* 290:1151-1155.

Lynch M and Walsh B. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates, Inc.

Lythgoe JN. 1979. *The ecology of vision*. Oxford: Clarendon Press.

Mackay TFC and Langley CH. 1990. Molecular and phenotypic variation in the *achaete-scute* region of *Drosophila melanogaster*. *Nature* 348:64-66.

Maddison WP and Maddison DR. *Mesquite: a modular system for evolutionary analysis, Version 1.12* 2001. Available from http://mesquiteproject.org.

Mann RS and Carroll SB. 2002. Molecular mechanisms of selector gene function and evolution. *Curr Opin Genet Dev* 12:592-600.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376-80.

Martins E. 2000. Adaptation and the comparative method. *Trends Ecol Evol* 15:296-299.

Martins E and Garland Jr. T. 1991. Phylogenetic analysis of the correlated evolution of continuous characters: a simulation study. *Evolution* 45:534-557.

McKaye KR and Marsh A. 1983. Food switching by two specialized algae-scraping cichlid fishes in Lake Malawi, Africa. *Oecologia* 56:245-248.

McKinney ML and McNamara KJ. 1991. *Heterochrony: the evolution of ontogeny*. New York: Plenum Press.

Mears AJ, Kondo M, Swain PK, Takada Y, Bush RA, Saunders TL, Sieving PA, and Swaroop A. 2001. Nrl is required for rod photoreceptor development. *Nat Genet* 28:447-452.

Meyer A, Kocher TD, Basasibwaki P, and Wilson AC. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* 347:550-3.

Meyer A and Wilson A. 1990. Origin of tetrapods inferred from thier mitochondrial DNA affiliation to lungfish. *Journal of Molecular Evolution* 31:359-364.

Midford PE, Garland T, and Maddison WP. *PDAP:PDTREE package for Mesquite* 2003. Available from http://mesquiteproject.org/pdap_mesquite.

Milam AH, Rose L, Cideciyan AV, Barakat MR, Tang W-X, Gupta N, Aleman TS, Wright AF, Stone EM, Sheffield VC, et al. 2002. The nuclear receptor NR2E3 plays a role in human retinal photoreceptor differentiation and degeneration. *Proc Natl Acad Sci U S A* 99:473-478.

Moran P and Kornfield I. 1993. Retention of an Ancestral Polymorphism in the Mbuna Species Flock (Teleostei: Cichlidae) of Lake Malawi. *Mol Biol Evol* 10:1015-1029.

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, and Cheung VG. 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature* 430:743-7.

Mundy NI. 2005. A window on the genetics of evolution: MC1R and plumage colouration in birds. *Proc R Soc B* 272:1633-1640.

Muntz WRA. 1976. Visual pigments of cichlid fishes from Malawi. *Vision Res* 16:897-903.

Nathans J, Thomas D, and Hogness DS. 1986. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* 232:193-202.

Neafsey DE and Hartl DL. 2005. Convergent loss of an anciently duplicated, functionally divergent RH2 opsin gene in the fugu and Tetraodon pufferfish lineages. *Gene* 350:161-71.

Néron B, Ménager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P, and Letondal C. 2009. Mobyle: a new full web bioinformatics framework. *Bioinformatics* 25:3005-3011.

Ng L, Hurley JB, Dierks B, Srinivas M, Salto C, Vennstrom B, Reh TA, and Forrest D. 2001. A thyroid hormone receptor that is required for the development of green cone photoreceptors. *Nat Genet* 27:94-98.

Novales-Flamarique H and Hawryshyn C. 1994. Ultraviolet Photoreception Contributes to Prey Search Behaviour in Two Species of Zooplanktivorous Fishes. *J Exp Biol* 186:187-98.

Nylander J, Ronquist F, Hulsenbeck JP, and Nieves-Aldrey J. 2004. Bayesian phylogenetic analysis of combinded data. *Syst Biol* 53:47-67.

O'Quin KE, Hofmann CM, Hofmann HA, and Carleton KL. 2010. Parallel evolution of opsin gene expression in African cichlid fishes. *Mol Biol Evol*:msq171.

O'Quin KE, Smith AR, Sharma A, and Carleton KL. 2011. New evidence for the role of heterochrony in the repeated evolution of cichlid opsin expression. *Evol Dev* in press.

Oakley TH. 2007. Today's multiple choice exam: (a) gene duplication; (b) structural mutation; (c) co-option; (d) regulatory mutation; (e) all of the above. *Evol Dev* 9:523-4.

Ohno S. 1970. *Evolution by Gene Duplication*. Berlin: Springer.

Olds LC and Sibley E. 2003. Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a cis regulatory element. *Hum Mol Genet* 12:2333-2340.

Oliver MK. 2011. *What is an utaka?*. Available from http://malawicichlids.com/mw01018.htm.

Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877-84.

Pagel M and Meade A. 2007. BayesTraits: Draft Manual. Reading, UK: School of Biological Sciences, University of Reading.

Pagel M, Meade A, and Barker D. 2004. Bayesian Estimation of Ancestral Character States on Phylogenies. *Syst Biol* 53:673-684.

Papalopulu N, Clarke JD, Bradley L, Wilkinson D, Krumlauf R, and Holder N. 1991. Retinoic acid causes abnormal development and segmental patterning of the anterior hindbrain in Xenopus embryos. *Development* 113:1145-1158.

Paradis E, Claude J, and Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20:289-290.

Parker A and Kornfield I. 1997. Evolution of the mitochondrial DNA control region in the mbuna (Cichlidae) species flock of Lake Malawi, East Africa. *J Mol Evol* 45:70-83.

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

Peng GH, Ahmad O, Ahmad F, Liu J, and Chen S. 2005. The photoreceptor-specific nuclear receptor Nr2e3 interacts with Crx and exerts opposing effects on the transcription of rod versus cone genes. *Hum Mol Genet* 14:747-64.

Pennacchio LA and Rubin EM. 2001. Genomic strategies to identify mammalian regulatory sequences. *Nat Rev Genet* 2:100-9.

Pollen AA, Dobberfuhl AP, Scace J, Igulu MM, Renn SC, Shumway CA, and Hofmann HA. 2007. Environmental complexity and social organization sculpt the brain in Lake Tanganyikan cichlid fish. *Brain Behav Evol* 70:21-39.

Posada D. 2008. jModelTest: Phylogenetic model averaging. *Mol Biol Evol* 25:1253-1256.

Posada D and Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817-8.

Price T. 2007. *Speciation in birds*. Greenwood Village, CO: Roberts & Co.

Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R, and Tabin CJ. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet* 38:107-11.

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

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559-75.

Purvis A and Garland Jr. T. 1993. Polytomies in comparative analyses of continuous characters. *Syst Biol* 42:569-575.

Quinn GP and Keough MJ. 2002. *Experimental design and data analysis for biologists*. New York, NY: Cambridge University Press.

R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing 2009. Available from http://www.R-project.org.

Raff RA. 1987. Constraint, flexibility, and phylogenetic history in the evolution of direct development in sea urchins. *Dev Biol* 19:6-19.

Ratkowsky RA. 1990. *Handbook of nonlinear regression models*. Edited by Owen DB. Vol. 107, *STATISTICS: Textbooks and monographs*. New York: Marcel Dekker, Inc.

Renn SCP, Aubin-Horth N, and Hofmann HA. 2008. Fish and chips: functional genomics of social plasticity in an African cichlid fish. *J Exp Biol* 211:3041-3056.

Repa JJ, Hanson KK, and Clagett-Dame M. 1993. All-trans-retinol is a ligand for the retinoic acid receptors. *Proc Natl Acad Sci U S A* 90:7293-7297.

Ribbink AJ, Marsh BA, Marsh AC, Ribbink AC, and Sharp BJ. 1983. A preliminary survey of the cichlid fishes of rocky habitats in Lake Malawi. *Suid-Afrikaanse Tydskrif vir Dierkunde* 18:149-310.

Roberts MR, Hendrickson A, McGuire CR, and Reh TA. 2005. Retinoid X Receptor γ Is Necessary to Establish the S-opsin Gradient in Cone Photoreceptors of the Developing Mouse Retina. *Invest Ophthalmol Vis Sci* 46:2897-2904.

Roberts RB, Ser JR, and Kocher TD. 2009. Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science* 326:998-1001.

Rost S, Fregin A, Ivaskevicius V, Conzelmann E, Hortnagel K, Pelz H-J, Lappegard K, Seifried E, Scharrer I, Tuddenham EGD, et al. 2004. Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature* 427:537-541.

Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, Bjornson R, Carriero N, Snyder M, and Gerstein MB. 2009. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat Biotechnol* 27:66-75.

Ryan DG, Oliveira-Fernandes M, and Lavker RM. 2006. MicroRNAs of the mammalian eye display distinct and overlaping tissue specificity. *Mol Vis* 12:1175-1184.

Salbert G, Fanjul A, Piedrafita FJ, Lu XP, Kim SJ, Tran P, and Pfahl M. 1993. Retinoic acid receptors and retinoid X receptor-alpha down-regulate the transforming growth factor-beta 1 promoter by antagonizing AP-1 activity. *Mol Endocrinol* 7:1347-56.

Salzburger W, Meyer A, Baric S, Verheyen E, and Sturmbauer C. 2002. Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. *Syst Biol* 51:113-35.

Schluter D. 1996. Adaptive Radiation Along Genetic Lines of Least Resistance. *Evolution* 50:1766-1774.

Schluter D. 2000. *The ecology of adaptive radiation*. Oxford, UK: Oxford University Press.

Schluter D, Price T, Mooers Aò, and Ludwig D. 1997. Likelihood of Ancestor States in Adaptive Radiation. *Evolution* 51:1699-1711.

Schug J. 2008. Using TESS to predict transcription factor binding sites in DNA sequence. *Curr Protoc Bioinformatics* Chapter 2:Unit 2 6.

Schule R, Rangarajan P, Yang N, Kliewer S, Ransone LJ, Bolado J, Verma IM, and Evans RM. 1991. Retinoic acid is a negative regulator of AP-1-responsive genes. *Proc Natl Acad Sci U S A* 88:6092-6.

Schwartz S, Elnitski L, Li M, Weirauch M, Riemer C, Smit A, Green ED, Hardison RC, and Miller W. 2003. MultiPipMaker and supporting tools: Alignments and analysis of multiple genomic DNA sequences. *Nucleic Acids Res* 31:3518-24.

Schwartz S, Zhang ZD, Smit A, Reimer C, Bouck C, Gibbs RA, Hardison RC, and Miller W. 2000. PipMaker--A web server for aligning two genomic DNA sequences. *Genome Res* 10:577-586.

Seehausen O. 2004. Hybridization and adaptive radiation. *Trends Ecol Evol* 19:198-207.

Seehausen O. 2006. African cichlid fish: a model system in adaptive radiation research. *Proc Biol Sci* 273:1987-98.

Seehausen O, Koetsier E, Schneider MV, Chapman LJ, Chapman CA, Knight ME, Turner GF, van Alphen JJ, and Bills R. 2003. Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. *Proc Biol Sci* 270:129-37.

Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HD, Miyagi R, van der Sluijs I, Schneider MV, Maan ME, Tachida H, et al. 2008. Speciation through sensory drive in cichlid fish. *Nature* 455:620-6.

Seehausen O, van Alphen JJ, and Witte F. 1997. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* 277:1808-1811.

Shand J, Davies WL, Thomas N, Balmer L, Cowing JA, Pointer M, Carvalho LS, Trezise AE, Collin SP, Beazley LD, et al. 2008. The influence of ontogeny and light environment on the expression of visual pigment opsins in the retina of the black bream, Acanthopagrus butcheri. *J Exp Biol* 211:1495-503.

Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schluter D, and Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428:717-23.

Siebeck UE and Marshall NJ. 2001. Ocular media transmission of coral reef fish--can coral reef fish see ultraviolet light? *Vision Res* 41:133-49.

Smallwood PM, Wang Y, and Nathans J. 2002. Role of a locus control region in the mutually exclusive expression of human red and green cone pigment genes. *Proc Natl Acad Sci U S A* 99:1008-11.

Smith A and Carleton K. 2010. Allelic Variation in Malawi Cichlid Opsins: A Tale of Two Genera. *J Mol Evol* 70:593-604.

Sokal RR and Rohlf FJ. 1995. *Biometry*. New York: W. H. Freeman and Company

Soler L, Conte M, Katagiri T, Howe A, Lee B-Y, Amemiya C, Stuart A, Dossat C, Poulain J, Johnson J, et al. 2010. Comparative physical maps derived from BAC end sequences of tilapia (Oreochromis niloticus). *BMC Genomics* 11:636.

Spady TC, Parry JW, Robinson PR, Hunt DM, Bowmaker JK, and Carleton KL. 2006. Evolution of the cichlid visual palette through ontogenetic subfunctionalization of the opsin gene arrays. *Mol Biol Evol* 23:1538-47.

Spady TC, Seehausen O, Loew ER, Jordan RC, Kocher TD, and Carleton KL. 2005. Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Mol Biol Evol* 22:1412-22.

Srinivas M, Ng L, Liu H, Jia L, and Forrest D. 2006. Activation of the Blue Opsin Gene in Cone Photoreceptor Development by Retinoid-Related Orphan Receptor {beta}. *Mol Endocrinol* 20:1728-1741.

Stark A, Brennecke J, Russell RB, and Cohen SM. 2003. Identification of Drosophila MicroRNA targets. *PLoS Biol* 1:E60.

Steinke D, Salzburger W, and Meyer A. 2006. Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. *J Mol Evol* 62:772-84.

Stern C and Tokunaga C. 1968. Autonomous pleiotropy in Drosophilia. *Proc Natl Acad Sci U S A* 60:1252-1259.

Stern DL. 2000. Evolutionary developmental biology and the problem of variation. *Evolution* 54:1079-1091.

Stern DL. 2006. Evolutionary developmental biology. In *Evolutionary Genetics: Concepts and Case Studies*. Oxford: Oxford University Press.

Stern DL and Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* 62:2155-77.

Stormo GD. 2000. DNA binding sites: representation and discovery. *Bioinformatics* 16:16-23.

Storz JF. 2005. INVITED REVIEW: Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* 14:671-688.

Streelman JT and Danley PD. 2003. The stages of vertebrate evolutionary radiation. *Trends Ecol Evol* 18:126-131.

Sturmbauer C. 1998. Explosive speciation in cichlid fishes of the African Great Lakes: a dynamic model of adaptive radiation. *Journal of Fish Biology* 53:18-36.

Sturmbauer C. 2005. Explosive speciation in cichlid fishes of the African Great Lakes: a dynamic model of adaptive radiation. *J Fish Biol* 53:18-36.

Sugawara T, Terai Y, Imai H, Turner GF, Koblmuller S, Sturmbauer C, Shichida Y, and Okada N. 2005. Parallelism of amino acid changes at the RH1 affecting spectral sensitivity among deep-water cichlids from Lakes Tanganyika and Malawi. *Proc Natl Acad Sci U S A* 102:5448-53.

Sugawara T, Terai Y, and Okada N. 2002. Natural selection of the rhodopsin gene during the adaptive radiation of East African Great Lakes cichlid fishes. *Mol Biol Evol* 19:1807-11.

Sung HM, Wang TY, Wang D, Huang YS, Wu JP, Tsai HK, Tzeng J, Huang CJ, Lee YC, Yang P, et al. 2009. Roles of trans and cis variation in yeast intraspecies evolution of gene expression. *Mol Biol Evol* 26:2533-8.

Swaroop A, Kim D, and Forrest D. 2010. Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. *Nat Rev Neurosci* 11:563-576.

Swofford DL. 2003. *Phylogenetic Analysis Using Parsimony (* and Other Methods) Version 4*. Sunderland: Sinauer Associates.

Taberlet P, Meyer A, and Bouvet J. 1992. Unusual mitochondrial DNA polymorphism in two local populations of blue tit Parus caeruleus. *Mol Ecol* 1:27-36.

Taborsky M, Hert E, Siemens MV, and Stoerig P. 1986. Social behaviour of *Lamprologus* species: functions and mechanisms. *Ann. Mus. R. Afr. Center Sci. Zool.* 251:7-11.

Takahashi K. 2003. Systematics of Tanganyikan cichlid fishes (Teleostei: Perciformes). *Ichthyiological Res* 50:367-382.

Takechi M, Seno S, and Kawamura S. 2008. Identification of cis-acting elements repressing blue opsin expression in zebrafish UV cones and pineal cells. *J Biol Chem* 283:31625-32.

Tamura K, Dudley J, Nei M, and Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-9.

Terai Y, Mayer WE, Klein J, Tichy H, and Okada N. 2002. The effect of selection on a long wavelength-sensitive (LWS) opsin gene of Lake Victoria cichlid fishes. *Proc Natl Acad Sci U S A* 99:15501-6.

Terai Y, Seehausen O, Sasaki T, Takahashi K, Mizoiri S, Sugawara T, Sato T, Watanabe M, Konijnendijk N, Mrosso HD, et al. 2006. Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biol* 4:e433.

Terakita A. 2005. The opsins. *Genome Biol* 6.

Tini M, Fraser RA, and Giguère V. 1995. Functional Interactions between Retinoic Acid Receptor-related Orphan Nuclear Receptor (ROR) and the Retinoic Acid Receptors in the Regulation of the F-Crystallin Promoter. *J Biol Chem* 270:20156-20161.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, et al. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 39:31-40.

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

Tsujimura T, Hosoya T, and Kawamura S. 2010. A Single Enhancer Regulating the Differential Expression of Duplicated Red-Sensitive Opsin Genes in Zebrafish. *PLoS Genet* 6:e1001245.

Tuan DY, Solomon WB, London IM, and Lee DP. 1989. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "beta-like globin" genes. *Proc Natl Acad Sci U S A* 86:2554-2558.

Umesono K, Murakami KK, Thompson CC, and Evans RM. 1991. Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors. *Cell* 65:1255-1266.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, and Leunissen JAM. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35:W71-W74.

Valentine JW. 1977. Transspecific evolution. In *Evolution*, edited by Dobzhansky T, Ayala FJ, Stebbins GL and Valentine JW. San Francisco: W. H. Freeman.

van der Meer HJ and Bowmaker JK. 1995. Interspecific variation of photoreceptors in four co-existing haplochromine cichlid fishes. *Brain Behav Evol* 45:232-240.

Wagner HJ and Kroger RH. 2000. Effects of long-term spectral deprivation on the morphological organization of the outer retina of the blue acara (Aequidens pulcher). *Philos Trans R Soc Lond B Biol Sci* 355:1249-52.

Wakefield MJ, Anderson M, Chang E, Wei KJ, Kaul R, Graves JA, Grutzner F, and Deeb SS. 2008. Cone visual pigments of monotremes: filling the phylogenetic gap. *Vis Neurosci* 25:257-64.

Wald G. 1935. Carotenoids and the Visual Cycle. *J Gen Physiol* 19:351-371.

Wald G. 1968. The molecular basis of visual excitation. *Nature* 219:800-7.

Wang Y, Macke JP, Merbs SL, Zack DJ, Klaunberg B, Bennett J, Gearhart J, and Nathans J. 1992. A locus control region adjacent to the human red and green visual pigment genes. *Neuron* 9:429-40.

Wasserman WW and Sandelin A. 2004. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet* 5:276-87.

Watson CT, Lubieniecki KP, Loew E, Davidson WS, and Breden F. 2010. Genomic organization of duplicated short wave-sensitive and long wave-sensitive opsin genes in the green swordtail, Xiphophorus helleri. *BMC Evol Biol* 10:87.

West-Eberhard MJ. 2003. *Developmental plasticity and evolution*. New York: Oxford University Press.

Wienholds E, Kloosterman WP, Miska E, Alvarez-Saavedra E, Berezikov E, de Bruijn E, Horvitz HR, Kauppinen S, and Plasterk RHA. 2005. MicroRNA Expression in Zebrafish Embryonic Development. *Science* 309:310-311.

Wittkopp PJ, Haerum BK, and Clark AG. 2004. Evolutionary changes in cis and trans gene regulation. *Nature* 430:85-8.

Wittkopp PJ, Williams BL, Selegue JE, and Carroll SB. 2003. Drosophila pigmentation evolution: divergent genotypes underlying convergent phenotypes. *Proc Natl Acad Sci U S A* 100:1808-13.

Won Y-J, Wang Y, Sivasundar A, Raincrow J, and Hey J. 2006. Nuclear gene variation and molecular dating of the cichlid species flock of Lake Malawi. *Mol Biol Evol* 23:828-837.

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 8:206-16.

Wray GA and McClay DR. 1989. Molecular heterochronies and heterotopies in early echinoid development. *Evolution* 43:803-813.

Xu S, Witmer PD, Lumayag S, Kovacs B, and Valle D. 2007. MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster. *J Biol Chem* 282:25053-66.

Yamaoka K, Hori M, and Kuratani S. 1986. Ecomorphology of feeding in "goby-like" cichlid fishes in Lake Tanganyika. *Physiology & Ecology Japan* 23:17-29.

Yokoyama R and Yokoyama S. 1990. Isolation, DNA sequence and evolution of a color visual pigment gene of the blind cave fish Astyanax fasciatus. *Vision Res* 30:807-816.

Yokoyama S. 2002. Molecular evolution of color vision in vertebrates. *Gene* 300:69-78.

Yokoyama S. 2008. Evolution of dim-light and color vision pigments. *Annu Rev Genomics Hum Genet* 9:259-82.

Yokoyama S and Tada T. 2010. Evolutionary dynamics of rhodopsin type 2 opsins in vertebrates. *Mol Biol Evol* 27:133-141.

Yokoyama S, Tada T, Zhang H, and Britt L. 2008. Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. *Proc Natl Acad Sci U S A* 105:13480-5.

Yokoyama S and Yokoyama R. 1996. Adaptive evolution of photoreceptors and visual pigments in vertebrates. *Annu Rev Ecol Syst* 27:543-567.

Yokoyama S, Zhang H, Radlwimmer FB, and Blow NS. 1999. Adaptive evolution of color vision of the Comoran coelacanth (Latimeria chalumnae). *Proc Natl Acad Sci U S A* 96:6279-84.

Yuh C-H, Bolouri H, and Davidson EH. 1998. Genomic Cis-Regulatory Logic: Experimental and Computational Analysis of a Sea Urchin Gene. *Science* 279:1896-1902.

Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, and Kruglyak L. 2003. *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet* 37:57-64.

Zhao JJ, Yang J, Lin J, Yao N, Zhu Y, Zheng J, Xu J, Cheng JQ, Lin JY, and Ma X. 2009. Identification of miRNAs associated with tumorigenesis of retinoblastoma by miRNA microarray analysis. *Childs Nerv Syst* 25:13-20.