The Molecular Basis of Colour Vision in Colourful Fish:
Four Long Wave-Sensitive (LWS) Opsins in Guppies (*Poecilia reticulata*) are Defined by Amino Acid Substitutions at Key Functional Sites

by

Matthew Norman Ward
B.Sc., University of Victoria, 2005

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In the Department of Biology

The Molecular Basis of Colour Vision in Colourful Fish: Four Long Wave-Sensitive (LWS) Opsins in Guppies (*Poecilia reticulata*) are Defined by Amino Acid Substitutions at Key Functional Sites

by

Matthew Norman Ward
B.Sc., University of Victoria, 2005

Supervisory Committee

Dr. John S. Taylor, Supervisor
(Department of Biology)

Dr. Robert Chow, Departmental Member
(Department of Biology)

Dr. Tom Reimchen, Departmental Member
(Department of Biology)

Supervisory Committee

Dr. John S. Taylor, Supervisor
(Department of Biology)

Dr. Robert Chow, Departmental Member
(Department of Biology)

Dr. Tom Reimchen, Departmental Member
(Department of Biology)

Abstract

Comparisons of functionally important changes at the molecular level in model systems have identified key adaptations driving isolation and speciation. For example, recent studies suggest mutations in opsins, the genes that encode retinal photoreceptors, influence colour-based female mate choice and have, thereby, contributed to the radiation of African cichlids. To test the hypothesis that the evolution of elaborate colouration in guppies (*Poecilia reticulata*) is also associated with opsin gene diversity, long wavelength-sensitive (LWS) opsin genes were sequenced in six species of the family Poeciliidae. Sequences of four LWS genes were amplified from the guppy genome and from mRNA isolated from adult eyes. Variation in expression was quantified using qPCR. Three of the four paralogs encode opsins that appear to be most sensitive to different wavelengths of light, as they vary at key amino acid positions. This family of LWS opsin genes was produced by a diversity of duplication events. One gene appears to be the product of retrotransposition. Between-gene PCR and DNA sequencing show that two of the guppy LWS opsins are linked in an inverted orientation. The fourth locus is a hybrid, with the first five exons most similar to one of the tandem duplicates and the last exon most similar to the other. The single-exon LWS opsin was produced prior to the divergence of families Fundulidae and Poeciliidae. The inverted tandem duplication event occurred near the base of the poecilid tree, in the common ancestor of *Poecilia* and *Xiphophorus,* and the hybrid locus is found only in the genus *Poecilia*. Enhanced wavelength discrimination, a likely consequence of opsin gene duplication and divergence, might have been an evolutionary prerequisite for colour based sexual selection and have led to the extraordinary colouration now observed in male guppies and in many other poeciliids.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

a.a.	Amino acid
ATP	Adenosine triphosphate
BfaI	Restriction enzyme from *Bacteroides fragilis*
BLAST	Basic local alignment search tool
BLASTp	Basic local alignment search tool for proteins
bp	Base-pair(s)
°C	Celsius
cDNA	Complementary deoxyribonucleic acid
CNS	Central Nervous System
dH_2O	Distilled water
DIG	Digoxigenin
DNA	Deoxyribonucleic acid
dNTP	Nucleotide triphosphates
E. coli	*Escherichia coli*
GPCR	G-protein coupled receptor
HF buffer	High fidelity buffer
hr	Hour(s)
ISH	*In situ* hybridization
Kbp	Kilo-base-pair(s)
LBA	Long branch attraction
LI-COR sequencer	Lambda instruments corporation sequencer
LWS	Long-wave sensitive
MEGA	Molecular evolution genetics analysis
mJ	Millijoule
mL	Milliliter
mm	Millimeter
mM	Millimolar
MMR	Mismatch repair
mRNA	Messenger ribonucleic acid
MP	Maximum parsimony
MPI	Message Passing Interface

MS222	Tricaine methanesulfonate
MSP	Microspectrophotometry
MWS	Middle-wave sensitive
NCBI	National Center for Biotechnology Information
ng	Nanogram
NJ	Neighbor joining
nm	Nanometer
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RSq	r-squared
RT-PCR	Reverse transcriptase polymerase chain reaction
SNP	Single nucleotide polymorphism
SSC	Sodium chloride sodium citrate
SWS	Short-wave sensitive
TE buffer	Tris-ethylenediaminetetraacetic acid buffer
UTR	Untranslated region
UV	Ultraviolet
XSrc	Tyrosine kinase control gene
3’	Three prime
5’	Five prime
µg	Microgram
µL	Microliter
μM	Micromolar
λ_{max}	Maximal wavelength sensitivity

Acknowledgements

Chapter 1 of this thesis is a paper prepared for publication entitled: The molecular basis of colour vision in colourful fish: "Four Long Wave-Sensitive (LWS) opsins in guppies (Poecilia reticulata) are defined by amino acid substitutions at key functional sites." This publication is currently in revision for the journal BMC Evolutionary Biology. The author list for this paper is: Matthew N. Ward, Allison M. Churcher, Kevin J. Dick, Chris R.J. Laver, Greg L. Owens, Megan D. Polack, Pam R. Ward, Felix Breden, and John S. Taylor. The first author is Matthew N. Ward and the corresponding authors are Felix Breden and John S. Taylor.

For chapter 1, John S. Taylor supervised the study. Matthew N. Ward and John S. Taylor designed and implemented wet-lab experimentation and data analysis. Allison M. Churcher carried out the template switching/mismatch repair experiment. Kevin J. Dick carried out the between-gene PCR and related sequencing experiments. Chris R.J. Laver and Greg L. Owens assisted with RT-PCR, qPCR and sequencing of the LWS S180r locus. Felix Breden and John S. Taylor obtained all fish samples. Megan D. Polack and Pam R. Ward assisted in RT-PCR. Matthew N. Ward and John S. Taylor wrote the manuscript with editing assistance from Felix Breden.

The authors of this manuscript (chapter 1) thank Roderick Haesevoets for help with DNA sequencing, Shelby Temple for comments on the manuscript and Gustavo Ybazeta for help on the phylogenetic analysis. The Environmental Protection Agency of Guyana, Ministry of Agriculture of Guyana, and Animal Husbandry and Fisheries of Suriname provided research and collecting permits. Fish were collected in Venezuela in collaboration with D. Taphorn of UNELLEZ (permit no. 0497). This work was funded by grants from the Canadian Foundation for Innovation and the British Columbia Knowledge Development Fund (John S. Taylor), Natural Sciences and Engineering Research Council of Canada (John S. Taylor and Felix Breden) and the National Geographic Society (Felix Breden).

For research guidance, editing assistance and departmental support with this thesis, I wish to thank my supervisor, John S. Taylor, and committee members, Tom Reimchen and Robert Chow. Without their strong influence and continuing support, this thesis would never have been possible.

Dedications

I dedicate this thesis to my loving family: my amazing finaceé Melissa, my wonderful parents Claire and David, my devoted brother Jamie, and my loving grandparents Norman, Nita and George. Without your continuing love, patience and support, this thesis would never have been possible.

I also wish to thank and dedicate this work to my mentor and supervisor, John Taylor. I can not thank you enough for all the time you have dedicated in helping me through this project.

Thesis Introduction

Colour detection represents an important ecological adaptation present in many species to facilitate numerous biological processes. For instance, predator-prey interactions, sexual selection, foraging efficiency and speciation events may in part, be each underscored by this important sensory mode. Of importance however, it has become clear that for a given species to detect and discriminate among colours; both colour vision and colour perception must exist (Jones et al., 2007). Colour vision relies on a species' visual system, which in turn, should be able to discriminate among different wavelengths of light and send this information to the CNS in an orderly fashion (Lennie, 2000; Jones et al., 2007). This process is facilitated on a molecular level by retinal cone cell photopigments. That is, protein opsins, which when linked to a chromophore (11-*cis*-retinal) can absorb photons of a specific wavelength range and initiate neural hyperpolarization for cascade signaling to the brain (Carroll, 2007). Colour perception however, is harder to define, as its underlying etiology manifests itself through an individual's biology, psychology and phenomenology (i.e. the subjective experience of an individual, which can not be understood by an external observer) (Lotto and Purves, 2002). In short however, colour perception could be thought of as the brain's ability to process the information acquired through colour vision, and interpret this in a meaningful way (Jones et al., 2007).

Intuitively, one might assume that for an individual to detect colour, both colour vision and colour perception would evolve simultaneously and in a reciprocal manner. In other words, the neural circuitry which facilitates colour perception would in turn, be linked to the colour vision capacity of that individual (i.e. the repertoire of expressed visual photopigments). However, this synergistic effect between these two phenomena may in fact only exist with regards to colour perception and not necessarily for colour vision. For instance, an individual who has become colour-blind later in life could still posses the perceptual capacity to sub-consciously see colour, yet could no longer detect colour because of the lack of colour vision. Conversely, female mice expressing a knock-in human LWS photopigment show enhanced long-wavelength sensitivity and a new capacity for acute red-orange colour discrimination (Jacobs et al., 2007). Indeed, through the mechanism of X-chromosome inactivation, human women with a four-photopigment genotype can perceive significantly more chromatic appearances than men and women who possess wild-type trichromatic vision (Jameson et al., 2001). Therefore in both mice and humans; colour vision appears to facilitate colour perception and in turn, allows for a greater ability to discriminate among colours. Further to this point, the genetic repertoire of photopigments which characterize the underlying basis of colour vision should consequently play an important role for determination of colour detection ability. This notion of colour detection ability among species will therefore represent an important assumption for this thesis.

The evolution, organization and functionality of colour vision have become important topics in molecular biology. Much insight into how species evolve, survive and co-exist in their surrounding environments has been gained through examination of the diverse mechanisms and underlying etiologies of this important sensory mode. Of importance, it has become clear that species which posses an expanded and diverged number of expressed visual opsins exhibit an increased spectral sensitivity to specific wavelengths of light (Briscoe, 2001; Jameson et al., 2001; Deeb, 2004). For instance, the butterfly (*Lycaena rubidus*), which has evolved sexually dimorphic eyes, possess diverged blue opsin duplicates allowing for more than a 63nm difference in maximal wavelength sensitivity for the short-wave spectral region (Sison-Mangus et al., 2006). Zebrafish (*Danio rerio*), which posses two LWS opsin genes with spectral sensitivities differing by 7nm, express their opsin paralogs in different portions of the retina and at different times during development (Raymond et al., 1995; Robinson et al., 1995; Chinen et al., 2003; Takechi and Kawamura, 2005). Finally, opsin duplication in New World primates has produced two X-linked MWS and LWS opsin genes, with trichormacy being reached in females who posses both alleles at one of these opsin paralogs (Dulai et al., 1999).

Although species that possess a diversity of opsin genes have the potential for increased wavelength discrimination; improved colour vision can only occur if the opsin paralogs are expressed simultaneously. Furthermore, photoreceptors located in the outermost layer of the retina typically exhibit a mosaic array (parcticularly in teleost fish), with distinct cone types occupying specific positions within the array (Cepko, 1996; Hoke *et al*, 2006; Kunz *et al*, 1983). To allow for discrimination of two different wavelengths of light innervating the same unit mosaic, yet detected maximally by a different duplicate opsin; both paralogs should be expressed spatially adjacent to one another. The advantage of evolving opsins that facilitate specific visual tuning appears to vary among species. Nonetheless, predator avoidance, sexual selection and prey-detection are all but a few of the common examples necessitating the need for species to evolve specialized spectral acuities. In the guppy (*Poecilia reticulata*), a similar story of evolved colour discrimination appears to have occurred. It is thus of great importance to carefully consider the modes, implications and underlying etiology that an enhanced wavelength sensitivity may have on this species and in its closely related family.

The guppy has been shown to exhibit extreme population genetic variation for both male colouration and female responsiveness through sexual selection (Hughes et al., 1999; Brooks, 2002). Specifically, male disparity in carotenoid colouration and red pterdine pigmentation (drosopterins) has been correlated with female mate-choice (Houde and Endler, 1990; Grether et al., 2001; Grether et al., 2005). This appears to be the result of a genetically associated sensory bias present in both sexes for an attraction to orange colouration, which has also been linked to copulation duration (Grether et al., 2005; Pilastro et al., 2007). As the expression of a costly trait can increase the chances of predation, the overall brightness and complexity of this ornamental pigmentation may send condition dependent signals of fitness towards potential female counterparts (Nicoletto, 1993; Godin and Dugatkin, 1996). Since males cannot synthesize carotenoid or pterdine pigments (which are obtained from dietary sources), carotenoid colouration may be a direct indicator of male foraging ability and as such, may be seen by females as an honest signal of mate quality (Grether et al., 2001). Therefore, male pigment ratios or hues in the guppy add not only to the polymorphic variation within and among different populations, but to female responsiveness and choosiness in selecting a mate for copulation (Brooks, 2002; Grether et al., 2001). Although the exact etiology and rationale behind this pre-existing preference has not yet been fully elucidated, its mechanism for selection must rely in the sensory mode of colour vision (and hence, colour detection). Thus for selection to occur based on male orange and red colouration in the highly variable populations of the guppy, it might be expected that long-wavelength sensitive colour vision in these fish would be highly fine tuned.

We have shown that the guppy, and its closely related species, posses four distinct long wavelength-sensitive (LWS) opsin paralogs; a greater complement of red-sensitive opsins than any other vertebrate. Three of these duplicates are defined by unique amino acid substitutions at Yokoyama and Radlwimmer's (2001) five key-sites for spectral tuning; likely facilitating three distinct visual sensitivity curves in the red/orange colour spectrum of these fish. Furthermore, all four LWS opsin paralogs were shown to be simultaneously expressed at different levels in the adult guppy retina. Microspectrophotometry (MSP) data from Archer et al. (1987 and 1990), have shown that guppies possess a polymorphic population of long-wavelength cone cells; absorbing light in the 529-579nm range. These results support the prediction that the LWS opsin key-site substitutions characterized in this study contribute to the observed increase in red/orange spectral sensitivity of these fish. Furthermore, individual guppies have also been shown to vary with respect to λ_{max} sensitivity within and between populations (Archer et al., 1987 and 1990). The expansion in cone cell diversity appears to be due to species-level gene duplication and divergence; with long-wavelength spectral variability possibly being controlled at both the genotype and expressional level. As such, this hypothesis could help to explain the wide range of female preference observed in the polymorphic populations of the guppy, and in turn, has perhaps facilitated male colour variation. It is therefore of great importance to properly characterize the evolution, functionality, implications and consequences of LWS opsin gene duplication and divergence in this model organism.

In this thesis, the characterization of LWS opsin paralogs in the guppy and other species of the family Poeciliidae is presented in two chapters. Chapter 1 is a publication currently in press by Ward et al. (2008) (see acknowledgements) entitled: "The molecular basis of colour vision in colourful fish: Four Long Wave-Sensitive (LWS) opsins in guppies (*Poecilia reticulata*) are defined by amino acid substitutions at key functional sites." Here, the characterization, evolution and implications of LWS opsin gene duplication and divergence in Poeciliidae are discussed. Chapter 2 of this thesis includes important work produced on this subject that was not included in the chapter 1 manuscript. Here, detailed discussions regarding PCR template switching, Poeciliidae LWS opsin alleles versus loci, analysis of work published by Hoffman et al. (2007) and Weadick and Chang (2007), investigation of LWS opsin paralog expression, and future directions, is presented.

Chapter 1

Research arcticle in press for BMC Evolutionary Biology

The molecular basis of colour vision in colourful fish: Four Long Wave-Sensitive (LWS) opsins in guppies (*Poecilia reticulata*) are defined by amino acid substitutions at key functional sites.

Abstract

Background: Comparisons of functionally important changes at the molecular level in model systems have identified key adaptations driving isolation and speciation. In cichlids, for example, long wavelength-sensitive (LWS) opsins appear to play an important role in mate choice and male colour variation within and among species. To test the hypothesis that the evolution of elaborate colouration in male guppies (*Poecilia reticulata*) is also associated with opsin gene diversity, we sequenced long wavelength-sensitive (LWS) opsin genes in six species of the family Poeciliidae.

Results: Sequences of four LWS genes were amplified from the guppy genome and from mRNA isolated from adult guppy eyes. Variation in expression was quantified using qPCR. Three of the four genes encode opsins predicted to be most sensitive to different wavelengths of light because they vary at key amino acid positions. This family of LWS opsin genes was produced by a diversity of duplication events. One of the guppy LWS opsins, an intronless gene, was produced prior to the divergence of families Fundulidae and Poeciliidae. Between-gene PCR and DNA sequencing show that two of the guppy LWS opsins are linked in an inverted orientation. This inverted tandem duplication event occurred near the base of the poecilid tree in the common ancestor of *Poecilia* and *Xiphophorus.* The fourth sequence has been uncovered only in the genus *Poecilia*. In the guppies surveyed here, this sequence is a hybrid, with the 5' end most similar to one of the tandem duplicates and the 3' end identical to the other.

Conclusion: Enhanced wavelength discrimination, a possible consequence of opsin gene duplication and divergence, might have been an evolutionary prerequisite for colour based sexual selection and have led to the extraordinary colouration now observed in male guppies and in many other poeciliids.

Background

Understanding the molecular basis of characters shaped by selection is a major goal of evolutionary genetics. Of parcticular interest are genes that encode conspicuous secondary sexual traits in males and the genes that influence female preference for such traits (Andersson, 1994). Among fish; sticklebacks (family Gasterosteidae), cichlids (family Cichlidae), and poeciliids, including the guppy (*Poecilia reticulata*) and swordtails (genus *Xiphophorus*), are the most important models for the study of sexual selection driven by female choice. In each of these taxa, female mate choice is influenced by male colouration and in each group, male colouration and female preference are influenced by genetic and environmental factors (Breden and Stoner, 1987; Bakker, 1993; Houde, 1997; Haesler and Seehausen, 2005; Salzburger et al., 2005).

Mapping studies designed to uncover genes responsible for species- and population-level colour variation in cichlids and sticklebacks are underway (Streelman et al., 2003) but to date none have been identified. While it is also the case that no female preference loci have been uncovered in fish, many cichlid species and some populations possess unique opsin genes that provide strong candidates. Indeed, the only DNA sequences that have been found to differ among the 200 to 500 endemic Lake Victoria haplochromine species are long wave-sensitive (LWS) opsins (Terai et al., 2002; Terai et al., 2006). In the cichlid genus *Pundamilia,* LWS opsin sequence and expression appears to be tuned to specific male colour morphs (Carlton et al., 2005). Thus, it appears that variation in LWS opsin genes influences female mate choice and speciation in this family (Maan et al., 2006).

Opsin genes encode membrane-bound receptors that are expressed primarily in rod and cone cells of the retina. Each opsin protein is associated with a chromophore and when exposed to light, this complex changes shape leading to rod or cone cell hyperpolarization (Masland, 2001). The detection of light at the receptor level requires input from just one type of opsin-chromophore receptor. However, discriminating among colours (wavelengths) involves the interpretation of signals from multiple adjacent retinal cone cells expressing different opsins. These different opsins often have names that reflect the wavelength of light that they are most sensitive to. For instance, short wave-sensitive (SWS), middle wave-sensitive (MWS), and long wave-sensitive opsins (LWS) are most sensitive to blue, green and red light, respectively. Gene duplication and divergence has generated this opsin diversity. For example, the human MWS opsin is a duplicate (or paralog) of the LWS opsin locus and now differs at three of the five amino acid positions known to influence wavelength sensitivity (Nathans et al., 1986; Dulai et al., 1999; Yokoyama and Radlwimmer, 1999; Yokoyama and Radlwimmer, 2001). Zebrafish also have a pair of LWS opsin genes with different five key-site haplotypes (Chinen et al., 2003).

The purpose of this study was to characterize LWS opsin gene sequence variation in guppies and in closely related species. We focused on this gene because microspectrophotometry (MSP) data indicated that guppies express more than one type of LWS opsin (Archer et al., 1987; Archer and Lythgoe, 1990) and because orange is an important component of female mate choice for these fish (Endler, 1980; Houde, 1997; Bourne et al., 2003; Körner et al., 2006). While two recent studies have reported LWS opsin gene variation in guppies (Hoffmann et al., 2007; Weadick and Chang, 2007) one focused only on short amplicons from a single fish and both presented incomplete data on the key-site amino acids known to influence spectral sensitivity. Genomic sequences, transcript expression levels, and data from other poeciliids have also not been reported to date.

We show that guppies (*Poecilia reticulata*) and three species in the guppy sister group (Micropoecilia) have four LWS genes. Sequence variation at the five key sites indicates that three of these LWS opsins are most sensitive to different wavelengths of light, providing *Poecilia* with a larger repertoire of LWS pigments than any other fish taxon. One of the guppy LWS opsins appears to be a single-exon gene, likely arising from a retrotransposition event. This gene was sequenced in all poeciliids surveyed and has been reported in the killifish, *Lucania goodie* (family Fundulidae). Two LWS opsins are linked, oriented in a tail-to-tail fashion, and separated by approximately 3.3 Kbp. The fourth is found only in the genus *Poecilia*. This is a hybrid or mosaic sequence in the guppies surveyed from Cumaná Venezuela. One LWS gene was amplified and sequenced from *Tomeurus gracilis* and three from the swordtail *Xiphophorus pygmaeus*. All four LWS opsins in the guppy were amplified from RNA isolated from adult eyes, but qPCR experiments show much variation among these duplicates in the level of expression.

Materials and methods

Genomic PCR and sequencing

All species surveyed are in the family Poeciliidae. Long wavelength-sensitive (LWS) opsin genes were amplified from DNA isolated from *Tomeurus gracilis* (one individual), *Xiphophorus pygmaeus* (one individual), and from four species in the genus *Poecilia*; *P. reticulata* (14 individuals), *P. picta* (four individuals) *P. parae* (three individuals) and *P. bifurca* (three individuals). The genus *Tomeurus* is the sister group to a clade that includes *Poecilia* and *Xiphophorus*, and most other poeciliids (Meyer and Lydeard, 1993; Hrbek et al., 2007). *Poecilia picta*, *Poecilia parae*, and *Poecilia bifurca*, occur in the sister taxon to the guppy (Breden et al., 1999). They were in a separate genus previous to Rosen and Bailey's (1963) revision of the poeciliids and we refer to them collectively as Micropoecilia. *Poecilia reticulata* (the guppy) was sampled from a population collected in Cumaná, Venezuela and bred in our laboratory aquarium. The Cumaná guppy has also been referred to as Endler's guppy, but is closely related to other guppy populations (Alexander and Breden, 2004). PCR reactions were run using genomic DNA isolated from fish euthanized with MS222 (Sigma® A5040) or from specimens preserved in 95% ethanol.

Initially, PCR and sequence data were obtained using primers ForBeg, Fw1a and Rev5, which are complementary to conserved regions of fish LWS opsin genes in exon I (ForBeg), exon II (Fw1a), and exon V (Rev5) (see tables 1.1 and 1.2). After uncovering multiple LWS opsin sequences in *Poecilia* (see results), we attempted to PCR-amplify

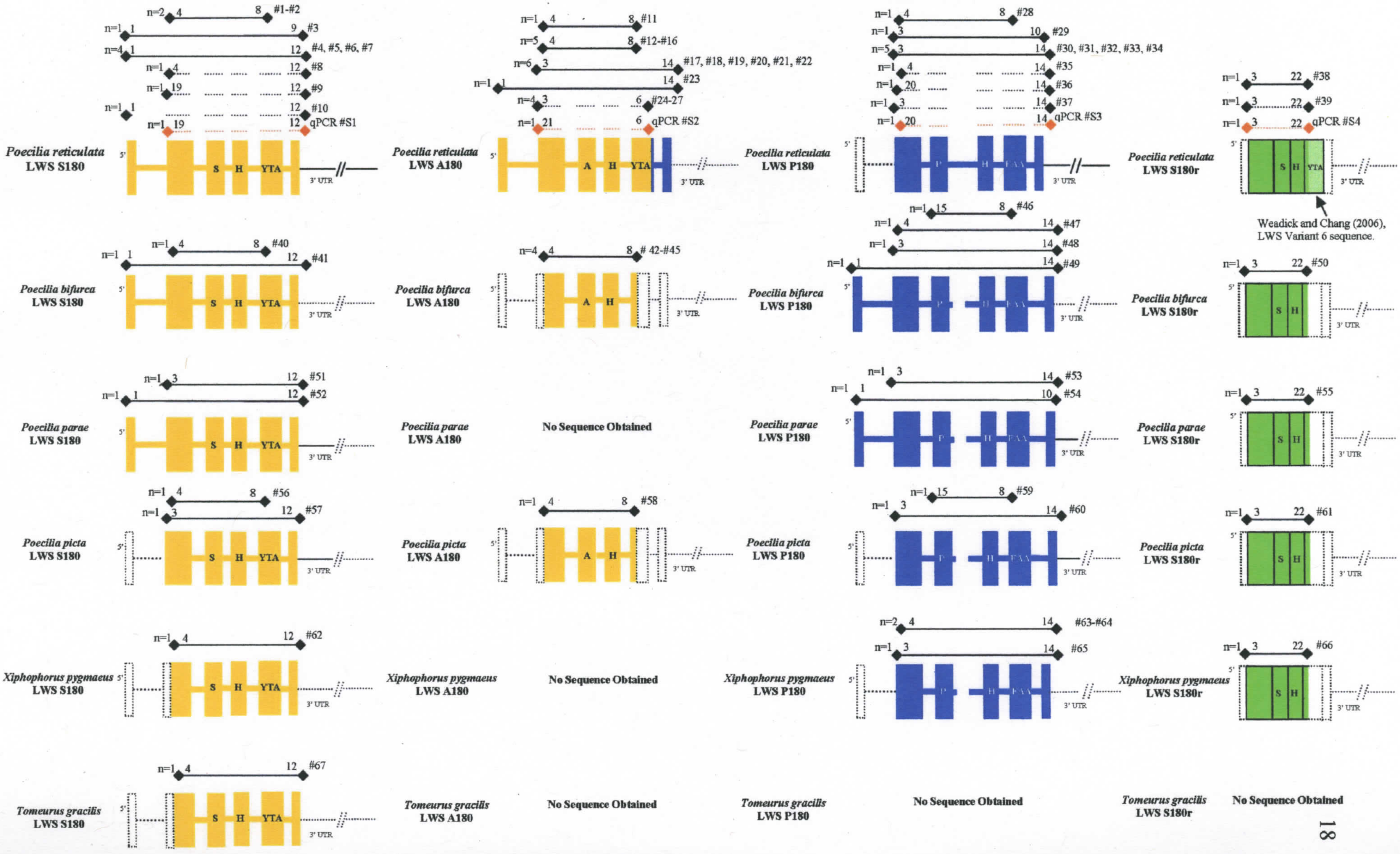
DNA between guppy opsin genes. The between-gene PCR experiment was initiated because LWS opsins occur in tandem in human, zebrafish, and medaka. It employed the reverse complement of a forward primer close to the 5' end of the gene (Fw1a Comp) and the reverse complement of a reverse primer close to the 3' end of the gene (Rev8 Comp). Sequence data from amplicons derived from primers ForBeg, Fw1a and Rev5, and the success of between-gene PCR allowed us to develop gene-specific primers, including reverse primers that were complementary to 3' UTR sequences. Primers complementary only to guppy LWS 'variant 6' were designed from sequence data recently published by Weadick and Chang (2007). PCR amplicons were cut and purified from agarose gels using a QIAquick® Gel Extraction Kit and were cloned using the pGEM® – T Easy Vector System II kit (Promega™). Sequencing of insert-positive clones utilized labeled M13 forward and reverse primers and a LI-COR sequencer at the Centre for Biomedical Research at the University of Victoria. A list of PCR primer sequences, PCR reaction conditions and genetic sequence maps can be found in table 1.1, table 1.2, and figure 1.1, respectively.

Primer Number	Primer Name	Sequence 5'→3'
1	ForBeg	ATGGCAGAGGAATGGGGAAAAC
2	ForEnd	TTATGCAGGAGCCACAGAGG
3	Fw100	GATCCCTTTGAAGGACCAAACT
4	Fw1a	TCTTATCAGTCTTCACCAACGG
5	Rev8 Comp	CTTCATGAACCGACAGGTGGGC
6	Rev8	GCCCACCTGTCGGTTCATGAAG
7	Rev4	GACCCAGGAGAAAACTATTCCAGC
8	Rev5	CATGACTACAACCATCCTGG
9	M13F	CAGTTTCTGTGCAGGTGACAGTAG
10	M13R	CGCTGATTGTTTATTCAGGTGC
11	LWS1 IntFor	GATAAACGGAAACTTTATGGCAAATG
12	LWS1 IntRev	CATTTGCCATAAAGTTTCCGTTTATC
13	LWS2 IntFor	GTTTGTTATTACTGCCGGGACTG
14	LWS2 IntRev	CAGTCCCGGCAGTAATAACAAAC
15	Long Intron F1	AGGAATTGCTGGGCTTTG
16	4KbFwdF	GCCTATTTTGATTTTATTGTTGATATTAC
17	4KbFwdR	CCTAGTGTCATCAGAAATACTAATCCATGTCCG
18	Guppy Gap	GAATTGTCTTGACTTGGGGTTGA
19	a/sExon2	GATGGGTTTACAACGTCTCCACAC
20	pExon2	GATGGGTTTACGATGTCGCAACGG
21	A180SpecExon2	GGGTTTACAACGTCTCCACTC
22	RevA	CATCCTAGATACTTCCTTCTGGG
23	Fw1a Comp	CCGTTGGTGAAGACTGATAAGA

Table 1.1. *List of LWS opsin primer names and sequences used for PCR, RT-PCR and qPCR*. Primer names and numbers correspond to reaction conditions shown in table 1.2. Primer numbers corresponds to amplicons shown in figure 1.1. Sequences are given in the 5'→ 3' orientation. Primers were synthesized by Operon® Biotechnologies, suspended in sterile buffered TE (pH 7.0) and stored for no longer than one year at -20°C.

Sequence Map Number	Species Name	LWS Name	Individual Name	Clone Name	Forward Primer	Reverse Primer	GenBank Identifier	Accession Number	PCR Conditions
#1	*Poecilia reticulata*	Genomic S180	All-11	1	(4) Fw1a	(8) Rev5	Guppy_LWS_S180_Seq#1-#2	EU329428	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#2	*Poecilia reticulata*	Genomic S180	All-11	2	(4) Fw1a	(8) Rev5	//	//	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#3	*Poecilia reticulata*	Genomic S180	GupF	-	(1) ForBeg	(9) M13F	Guppy_LWS_S180_Seq#3	EU329429	[94°C(10s) + 63°C(25s) + 72°C(50s)] x 35
#4	*Poecilia reticulata*	Genomic S180	4	-	(1) ForBeg	(12) LWS1 IntRev	Guppy_LWS_S180_Seq#4	EU329430	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#5	*Poecilia reticulata*	Genomic S180	5	-	(1) ForBeg	(12) LWS1 IntRev	Guppy_LWS_S180_Seq#5	EU329431	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#6	*Poecilia reticulata*	Genomic S180	6	-	(1) ForBeg	(12) LWS1 IntRev	Guppy_LWS_S180_Seq#6	EU329432	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#7	*Poecilia reticulata*	Genomic S180	10	-	(1) ForBeg	(12) LWS1 IntRev	Guppy_LWS_S180_Seq#7	EU329433	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#8	*Poecilia reticulata*	cDNA S180	Adult 1	-	(4) Fw1a	(12) LWS1 IntRev	Guppy_LWS_S180_Seq#8	EU329434	[94°C (6s) + 55°C(12s) + 72°C(21s)] x 35
#9	*Poecilia reticulata*	cDNA S180	Embryo 1	-	(19) a/sExon2	(12) LWS1 IntRev	Guppy_LWS_S180_Seq#9	EU329435	[94°C(10s) + 61°C(17s) + 72°C(27s)] x 35
#10	*Poecilia reticulata*	cDNA S180	Adult 2	-	(1) ForBeg	(12) LWS1 IntRev	Guppy_LWS_S180_Seq#10	EU329436	[94°C (8s) + 55°C(13s) + 72°C(25s)] x 35
#11	*Poecilia reticulata*	Genomic A180	d31	-	(4) Fw1a	(8) Rev5	Guppy_LWS_A180_Seq#11	EU329437	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#12	*Poecilia reticulata*	Genomic A180	Allison	1	(4) Fw1a	(8) Rev5	Guppy_LWS_A180_Seq#12-#16	EU329438	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#13	*Poecilia reticulata*	Genomic A180	Allison	2	(4) Fw1a	(8) Rev5	//	//	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#14	*Poecilia reticulata*	Genomic A180	Allison	3	(4) Fw1a	(8) Rev5	//	//	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#15	*Poecilia reticulata*	Genomic A180	Allison	4	(4) Fw1a	(8) Rev5	//	//	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#16	*Poecilia reticulata*	Genomic A180	Allison	5	(4) Fw1a	(8) Rev5	//	//	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#17	*Poecilia reticulata*	Genomic A180	6	-	(3) Fw100	(14) LWS2 IntRev	Guppy_LWS_A180_Seq#17	EU329439	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#18	*Poecilia reticulata*	Genomic A180	4	-	(3) Fw100	(14) LWS2 IntRev	Guppy_LWS_A180_Seq#18	EU329440	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#19	*Poecilia reticulata*	Genomic A180	Pam	-	(3) Fw100	(14) LWS2 IntRev	Guppy_LWS_A180_Seq#19	EU329441	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#20	*Poecilia reticulata*	Genomic A180	10	-	(3) Fw100	(14) LWS2 IntRev	Guppy_LWS_A180_Seq#20	EU329442	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#21	*Poecilia reticulata*	Genomic A180	5	-	(3) Fw100	(14) LWS2 IntRev	Guppy_LWS_A180_Seq#21	EU329443	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#22	*Poecilia reticulata*	Genomic A180	4	-	(3) Fw100	(14) LWS2 IntRev	Guppy_LWS_A180_Seq#22	EU329444	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#23	*Poecilia reticulata*	Genomic A180	6	-	(1) ForBeg	(14) LWS2 IntRev	Guppy_LWS_A180_Seq#23	EU329445	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#24	*Poecilia reticulata*	cDNA A180	Adult 1	1	(3) Fw100	(6) Rev8	Guppy_LWS_A180_Seq#24-#27	EU329446	[98°C (7s) + 60°C(20s) + 72°C(28s)] x 35
#25	*Poecilia reticulata*	cDNA A180	Adult 1	2	(3) Fw100	(6) Rev8	//	//	[98°C (7s) + 60°C(20s) + 72°C(28s)] x 35
#26	*Poecilia reticulata*	cDNA A180	Adult 1	3	(3) Fw100	(6) Rev8	//	//	[98°C (7s) + 60°C(20s) + 72°C(28s)] x 35
#27	*Poecilia reticulata*	cDNA A180	Adult 1	4	(3) Fw100	(6) Rev8	//	//	[98°C (7s) + 60°C(20s) + 72°C(28s)] x 35
#28	*Poecilia reticulata*	Genomic P180	Gupfb	-	(4) Fw1a	(8) Rev5	Guppy_LWS_P180_Seq#28	EU329447	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#29	*Poecilia reticulata*	Genomic P180	Gup100A	-	(3) Fw100	(10) M13R	Guppy_LWS_P180_Seq#29	EU329448	[94°C(10s) + 63°C(25s) + 72°C(45s)] x 35
#30	*Poecilia reticulata*	Genomic P180	4	-	(3) Fw100	(14) LWS2 IntRev	Guppy_LWS_P180_Seq#30	EU329449	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#31	*Poecilia reticulata*	Genomic P180	7	-	(3) Fw100	(14) LWS2 IntRev	Guppy_LWS_P180_Seq#31	EU329450	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#32	*Poecilia reticulata*	Genomic P180	8	-	(3) Fw100	(14) LWS2 IntRev	Guppy_LWS_P180_Seq#32	EU329451	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#33	*Poecilia reticulata*	Genomic P180	10	-	(3) Fw100	(14) LWS2 IntRev	Guppy_LWS_P180_Seq#33	EU329452	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#34	*Poecilia reticulata*	Genomic P180	Pam	-	(3) Fw100	(14) LWS2 IntRev	Guppy_LWS_P180_Seq#34	EU329453	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#35	*Poecilia reticulata*	cDNA P180	Adult 1	-	(4) Fw1a	(14) LWS2 IntRev	Guppy_LWS_P180_Seq#35	EU329454	[94°C (6s) + 54°C(12s) + 72°C(21s)] x 35
#36	*Poecilia reticulata*	cDNA P180	Embryo 1	-	(20) pExon2	(14) LWS2 IntRev	Guppy_LWS_P180_Seq#36	EU329455	[94°C(10s) + 55°C(17s) + 72°C(27s)] x 35
#37	*Poecilia reticulata*	cDNA P180	Adult 2	-	(3) Fw100	(14) LWS2 IntRev	Guppy_LWS_P180_Seq#37	EU329456	[94°C (6s) + 50°C(12s) + 72°C(21s)] x 35
#38	*Poecilia reticulata*	Genomic S180r	NWA	-	(3) Fw100	(22) RevA	Guppy_LWS_S180r_Seq#38	EU329457	[98°C (8s) + 60°C(20s) + 72°C(30s)] x 35
#39	*Poecilia reticulata*	cDNA S180r	JuvQ	-	(3) Fw100	(22) RevA	Guppy_LWS_S180r_Seq#39	EU329458	[98°C (8s) + 60°C(20s) + 72°C(30s)] x 35
#40	*Poecilia bifurca*	Genomic S180	2	-	(4) Fw1a	(8) Rev5	Bifurca_LWS_S180_Seq#40	EU329459	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#41	*Poecilia bifurca*	Genomic S180	1	-	(1) ForBeg	(12) LWS1 IntRev	Bifurca_LWS_S180_Seq#41	EU329460	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#42	*Poecilia bifurca*	Genomic A180	2	1	(4) Fw1a	(8) Rev5	Bifurca_LWS_A180_Seq#42-#45	EU329461	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#43	*Poecilia bifurca*	Genomic A180	2	2	(4) Fw1a	(8) Rev5	//	//	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#44	*Poecilia bifurca*	Genomic A180	2	3	(4) Fw1a	(8) Rev5	//	//	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#45	*Poecilia bifurca*	Genomic A180	2	4	(4) Fw1a	(8) Rev5	//	//	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#46	*Poecilia bifurca*	Genomic P180	1	-	(15) Long Intron F1	(8) Rev5	Bifurca_LWS_P180_Seq#46	EU329462	[94°C(10s) + 66°C(13s) + 72°C(25s)] x 35
#47	*Poecilia bifurca*	Genomic P180	3	-	(4) Fw1a	(14) LWS2 IntRev	Bifurca_LWS_P180_Seq#47	EU329463	[94°C(10s) + 59°C(25s) + 72°C(60s)] x 35
#48	*Poecilia bifurca*	Genomic P180	3	-	(3) Fw100	(14) LWS2 IntRev	Bifurca_LWS_P180_Seq#48	EU329464	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#49	*Poecilia bifurca*	Genomic P180	1	-	(1) ForBeg	(14) LWS2 IntRev	Bifurca_LWS_P180_Seq#49	EU329465	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#50	*Poecilia bifurca*	Genomic S180r	3	-	(3) Fw100	(22) RevA	Bifurca_LWS_S180r_Seq#50	EU329466	[98°C (8s) + 60°C(20s) + 72°C(30s)] x 35
#51	*Poecilia parae*	Genomic S180	1	-	(3) Fw100	(12) LWS1 IntRev	Parae_LWS_S180_Seq#51	EU329467	[94°C(10s) + 60°C(25s) + 72°C(55s)] x 35
#52	*Poecilia parae*	Genomic S180	1	-	(1) ForBeg	(12) LWS1 IntRev	Parae_LWS_S180_Seq#52	EU329468	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#53	*Poecilia parae*	Genomic P180	1	-	(3) Fw100	(14) LWS2 IntRev	Parae_LWS_P180_Seq#53	EU329469	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#54	*Poecilia parae*	Genomic P180	1	-	(1) ForBeg	(10) M13R	Parae_LWS_P180_Seq#54	EU329470	[94°C(10s) + 63°C(25s) + 72°C(50s)] x 35
#55	*Poecilia parae*	Genomic S180r	2	-	(3) Fw100	(22) RevA	Parae_LWS_S180r_Seq#55	EU329471	[98°C (8s) + 60°C(20s) + 72°C(30s)] x 35
#56	*Poecilia picta*	Genomic S180	2	-	(4) Fw1a	(8) Rev5	Picta_LWS_S180_Seq#56	EU329472	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#57	*Poecilia picta*	Genomic S180	1	-	(3) Fw100	(12) LWS1 IntRev	Picta_LWS_S180_Seq#57	EU329473	[94°C(10s) + 60°C(25s) + 72°C(55s)] x 35
#58	*Poecilia picta*	Genomic A180	2	-	(4) Fw1a	(8) Rev5	Picta_LWS_A180_Seq#58	EU329474	[94°C(10s) + 60°C(15s) + 72°C(25s)] x 35
#59	*Poecilia picta*	Genomic P180	2	-	(15) Long Intron F1	(8) Rev5	Picta_LWS_P180_Seq#59	EU329475	[94°C(10s) + 66°C(13s) + 72°C(25s)] x 35
#60	*Poecilia picta*	Genomic P180	1	-	(3) Fw100	(14) LWS2 IntRev	Picta_LWS_P180_Seq#60	EU329476	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#61	*Poecilia picta*	Genomic S180r	3	-	(3) Fw100	(22) RevA	Picta_LWS_S180r_Seq#61	EU329477	[98°C (8s) + 60°C(20s) + 72°C(30s)] x 35
#62	*Xiphophorus pyg.*	Genomic S180	1	-	(4) Fw1a	(12) LWS1 IntRev	Xiphophorus_LWS_S180_Seq#62	EU329478	[94°C(10s) + 59°C(25s) + 72°C(60s)] x 35
#63	*Xiphophorus pyg.*	Genomic P180	1	1	(4) Fw1a	(14) LWS2 IntRev	Xiphophorus_LWS_S180_Seq#63-#64	EU329479	[94°C(10s) + 59°C(25s) + 72°C(60s)] x 35
#64	*Xiphophorus pyg.*	Genomic P180	1	2	(4) Fw1a	(14) LWS2 IntRev	//	//	[94°C(10s) + 59°C(25s) + 72°C(60s)] x 35
#65	*Xiphophorus pyg.*	Genomic P180	1	-	(3) Fw100	(14) LWS2 IntRev	Xiphophorus_LWS_S180_Seq#65	EU329480	[94°C(10s) + 59°C(25s) + 72°C(65s)] x 35
#66	*Xiphophorus pyg.*	Genomic S180r	1	-	(3) Fw100	(22) RevA	Xiphophorus_LWS_S180_Seq#66	EU329481	[98°C (8s) + 60°C(20s) + 72°C(30s)] x 35
#67	*Tomeurus gracilis*	Genomic S180	1	-	(4) Fw1a	(12) LWS1 IntRev	Tomeurus_LWS_S180_Seq#67	EU329482	[94°C(10s) + 59°C(25s) + 72°C(60s)] x 35
qPCR #S1	*Poecilia reticulata*	cDNA S180	Juv1	-	(19) a/sExon2	(12) LWS1 IntRev	Guppy_LWS_S180_qPCRS1	EU329483	[98°C (8s) + 60°C(20s) + 72°C(30s)] x 35
qPCR #S2	*Poecilia reticulata*	cDNA A180	Juv1	-	(21) A180SpecExon2	(6) Rev8	Guppy_LWS_A180_qPCRS2	EU329484	[98°C (8s) + 60°C(20s) + 72°C(30s)] x 35
qPCR #S3	*Poecilia reticulata*	cDNA P180	Juv1	-	(20) pExon2	(14) LWS2 IntRev	Guppy_LWS_P180_qPCRS3	EU329485	[98°C (8s) + 60°C(20s) + 72°C(30s)] x 35
qPCR #S4	*Poecilia reticulata*	cDNA S180r	Juv1	-	(3) Fw100	(22) RevA	Guppy_LWS_S180r_qPCRS4	EU329486	[98°C (8s) + 60°C(20s) + 72°C(30s)] x 35
-	*Poecilia reticulata*	Intergenic S180 and P180	Gup4Kb	-	(5) Rev8 Comp	(5) Rev8 Comp	-	-	[94°C(10s) + 60°C(20s) + 72°C(95s)] x 35
-	*Poecilia reticulata*	†Intergenic S180 and P180	Gup4Kb	-	(11) LWS1 IntFor	(13) LWS2 IntFor	-	-	[94°C(10s) + 59°C(20s) + 72°C(85s)] x 35
-	*Poecilia reticulata*	†Intergenic S180 and P180	Gup4Kb	-	(16) 4KbFwdF	-	-	-	Direct sequencing from † clone
-	*Poecilia reticulata*	†Intergenic S180 and P180	Gup4Kb	-	(17) 4KbFwdR	-	-	-	Direct sequencing from † clone
-	*Poecilia reticulata*	†Intergenic S180 and P180	Gup4Kb	-	(18) Guppy Gap	-	-	-	Direct sequencing from † clone
-	*Poecilia parae*	Intergenic S180 and P180	Par4Kb	-	(5) Rev8 Comp	(5) Rev8 Comp	-	-	[94°C(10s) + 60°C(20s) + 72°C(95s)] x 35
-	*Poecilia picta*	Intergenic S180 and P180	Pic4Kb	-	(5) Rev8 Comp	(5) Rev8 Comp	-	-	[94°C(10s) + 60°C(20s) + 72°C(95s)] x 35

Table 1.2. *Sequence data and PCR conditions for specific primer combinations used in all six species of Poeciliidae.* Primer names are given with corresponding primer number from table 1.1. Amplicons are shown in figure 1.1 along with corresponding n value. 0.5U iProof™ DNA polymerase (BioRad®) was used for each reaction with 5X iProof™ HF Buffer, 10mM dNTP mix, ~100ng template DNA, 0.5μM of both forward and reverse primers and dH_2O. An Eppendorf® silver block thermal cycler was used for all PCR reactions. Each reaction included an initial denaturation of 94°C for 30 seconds and a final extension of 72°C for 600 seconds. If >1 band was found after gel electrophoresis, expected sized amplicons were cut and purified using a QIAquick® Gel Extraction Kit. †Cloned using a pGEM®-T easy Vector System II kit (Promega).§ Sequenced off of the † clone using a dye terminator cycle sequencer on a Beckman Coulter CEQ 8000 genetic analysis system (160 minutes) located at the University of Victoria.


Poecilia reticulata LWS S180
n=2 4 8 #1-#2
n=1 1 9 #3
n=4 1 12 #4, #5, #6, #7
n=1 4 12 #8
n=1 19 12 #9
n=1 1 12 #10
n=1 19 12 qPCR #S1
5'
S H YTA
3' UTR
Poecilia reticulata LWS A180
n=1 4 8 #11
n=5 4 8 #12-#16
n=6 3 14 #17, #18, #19, #20, #21, #22
n=1 1 14 #23
n=4 3 6 #24-27
n=1 21 6 qPCR #S2
A H YTA
Poecilia reticulata LWS P180
n=1 4 8 #28
n=1 3 10 #29
n=5 3 14 #30, #31, #32, #33, #34
n=1 4 14 #35
n=1 20 14 #36
n=1 3 14 #37
n=1 20 14 qPCR #S3
P H FAA
Poecilia reticulata LWS S180r
n=1 3 22 #38
n=1 3 22 #39
n=1 3 22 qPCR #S4
S H YTA
Weadick and Chang (2006), LWS Variant 6 sequence.
Poecilia bifurca LWS S180
n=1 4 8 #40
n=1 1 12 #41
Poecilia bifurca LWS A180
n=4 4 8 # 42-#45
Poecilia bifurca LWS P180
n=1 15 8 #46
n=1 4 14 #47
n=1 3 14 #48
n=1 1 14 #49
Poecilia bifurca LWS S180r
n=1 3 22 #50
Poecilia parae LWS S180
n=1 3 12 #51
n=1 1 12 #52
Poecilia parae LWS A180
No Sequence Obtained
Poecilia parae LWS P180
n=1 3 14 #53
n=1 1 10 #54
Poecilia parae LWS S180r
n=1 3 22 #55
Poecilia picta LWS S180
n=1 4 8 #56
n=1 3 12 #57
Poecilia picta LWS A180
n=1 4 8 #58
Poecilia picta LWS P180
n=1 15 8 #59
n=1 3 14 #60
Poecilia picta LWS S180r
n=1 3 22 #61
Xiphophorus pygmaeus LWS S180
n=1 4 12 #62
Xiphophorus pygmaeus LWS A180
No Sequence Obtained
Xiphophorus pygmaeus LWS P180
n=2 4 14 #63-#64
n=1 3 14 #65
Xiphophorus pygmaeus LWS S180r
n=1 3 22 #66
Tomeurus gracilis LWS S180
n=1 4 12 #67
Tomeurus gracilis LWS A180
No Sequence Obtained
Tomeurus gracilis LWS P180
No Sequence Obtained
Tomeurus gracilis LWS S180r
No Sequence Obtained

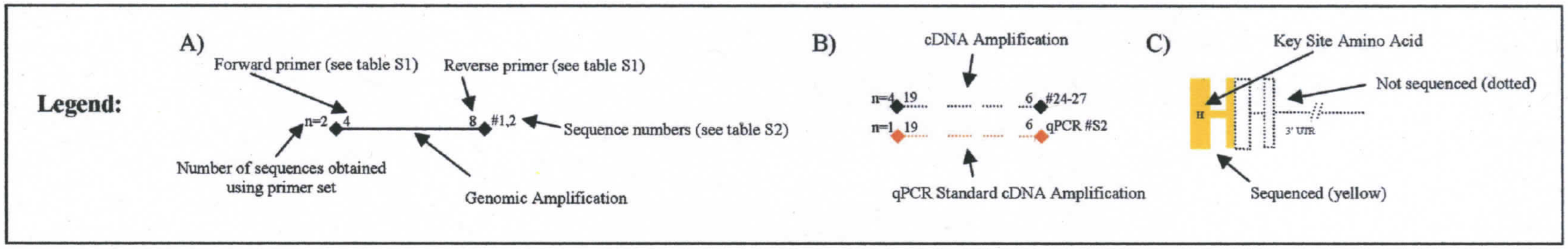


Figure 1.1. *Primer and sequence maps for LWS opsin loci in Poeciliidae.* Large coloured boxes and horizontal bars represent exons and introns, respectively. Black and solid horizontal lines represent genomic sequences amplified using various PCR primer combinations (see legend; see tables 1.1 and 1.2). Dotted horizontal lines represent mRNA transcripts acquired through RT-PCR from cDNA samples. "n" represents the number of times a given sequence of near perfect identity had been acquired. Key site amino acids are shown within their respective exons. Inter-species and individual sequence differences due to SNP's, small strings of nucleotides, or codon selection (see discussion) are omitted and a consensus map has been created. Yellow exons and introns are highly similar to the guppy LWS S180 locus. Blue colouration represents sequences of high identity to the guppy LWS P180 locus. This use of colour is exploited to highlight the hybrid nature of the LWS A180 gene in the guppy. Finally, green colouration represents the single exon LWS S180r locus, which was originally discovered by Weadick and Chang (2007) named 'variant 6' in their study.

Southern Blot Hybridization

DNA from one lab-reared Cumaná guppy was extracted using a Qiagen® DNeasy Tissue Kit and digested with the four-cutter restriction enzyme *Bfa*I (New England Biolabs®). A 423 bp DIG labeled probe was prepared from guppy genomic DNA using the PCR DIG Probe Synthesis Kit (Roche®) and primers Fw100 and Rev4 (see table 1.1). The probe was purified using the QIAquick® Gel Extraction Kit (Qiagen®). *Bfa*I (from *Bacteroides fragilis)* does not cut within a region of the LWS opsins that is complementary to the probe. Southern blot hybridization was carried out using a modified protocol from the Roche® DIG application manual for filter hybridization. Digested DNA was blotted onto a Bio-Rad™ Zeta-Probe® Blotting membrane using the Bio-Rad™ Model 785 Vacuum Blotter and was followed by UV exposure (120mJ) and probe hybridization at 40°C overnight. The blot was washed in 2X SSC at room temperature and then in 1X SSC at 65°C and visualized using the DIG Luminescent Detection Kit for Nucleic Acids (Roche®).

LWS opsin gene expression using RT-PCR

Prior to quantitative PCR (qPCR) experiments (see below) we tested the hypothesis that all four LWS opsin loci were expressed using reverse transcriptase (RT)-PCR. Three guppies were euthanized in MS222. A single eye from each individual was placed in 1.0 mL PureZOL™ (Bio-Rad®) with 3mm tungsten carbide beads and homogenized for five minutes in a Retsch MM301 Mixer Mill. Total RNA was extracted using the Aurum™ Total RNA Fatty and Fibrous Tissue kit from BioRad®. The iScript™ kit (Bio-Rad®) was used to generate single-stranded cDNA. LWS transcripts were PCR amplified and cloned using the pGEM® – T Easy Vector System II kit (Promega™), and sequenced with labeled M13 primers. Primers reported by Meyer and Lydeard (1993) that amplify *XSrc* were used as a positive control. Primer sequences and PCR conditions can be found in tables 1.1 and 1.2.

Quantitative PCR

Data from cichlids and zebrafish indicate that cone opsin expression is highest at dusk (Halstenberg et al., 2005; Li et al., 2005). Guppies in our lab were maintained in a 14:10 hr light and dark cycle and qPCR experiments were performed on cDNA samples obtained from three fish in the last hour of the subjective day. Total RNA was extracted from the eyes of these fish (one adult male and two adult females) using the Retsch MM301 Mixer Mill and the Aurum™ Total RNA Fatty and Fibrous Tissue kit from BioRad®. Synthesis of cDNA for qPCR experiments utilized the SuperScript™ III First-Strand Synthesis SuperMix kit for qRT-PCR (Invitrogen™) and 1 μg of total RNA from the three samples. To determine the concentration of each transcript in the three cDNA samples, we used the Invitrogen™ SYBR® GreenER™ qPCR SuperMix Universal kit to prepare triplet qPCR reactions. qPCR was carried out in a Stratagene® Mx4000® Multiplex Quantitative PCR machine with the following locus-specific primer pairs: a/sExon2 and LWS1IntRev; A180SpecFwd and rev8; pExon2 and LWS2IntRev; and fw100 plus revA (see table 1.1). A 1:10 ROX Reference Dye normalized the fluorescent reporter signal. qPCR conditions consisted of 1 cycle at 95°C (9 minutes); 50 cycles of 95°C (15 seconds), 60°C (30 seconds), 72°C (45 seconds); 1 cycle of 95°C (1 minute); and a 40-step melting curve analysis (initial temperature 55°C, increasing 1°C every 30 seconds). Each gene was also PCR amplified, cloned using the pGEM® – T Easy Vector

System II kit (Promega™), sequenced to confirm identity, and then utilized for qPCR at concentrations of 1ng, $1\text{x}10^{-3}$ng and $1\text{x}10^{-5}$ng per 16 µL reaction. Ct values from these plasmid templates were then used to generate a standard curve and to estimate qPCR efficiency ($\text{qPCR}_{\text{eff}} = [10^{(-1/\text{slope})} - 1] \text{ x } 100$; see figure 1.4). The plasmid template reactions were also run in triplicate. Dissociation curves (Fluorescence [–R'(T)] over T°C) and gel electrophoresis confirmed the presence of single amplicons in all qPCR reactions.

Phylogenetic analyses

Tetraodon nigroviridis LWS amino acid sequence (AAT38457.1) was employed as a query sequence in a BLASTp search (Altschul et al., 1990) to identify LWS homologs in the NCBI nr database. All hits with bitscores >300 were aligned with one another and with the new data using the MPI version of ClustalW (Thompson et al., 1994; Li, 2003). Subsequent sequence manipulations including multiple sequence alignments, toggle translations, hand editing, and delimitation of intron/exon boundaries utilized BioEdit v.7.0.5.3. (Hall, 1999). A short multiple sequence alignment (390bp) that included coding sequences from exons IV and V was used to determine relationships among our new guppy LWS opsin genes and those from guppies of the Oropuche and Quare Rivers in Trinidad reported by Hoffmann et al. (2007): OR6-4 D09/DQ168660.1 and OR6-3 EO8/DQ168659.1 and QUEm5 LO6/DQ168661.1, and from the Paria River in Trinidad reported by Weadick and Chang (2007): DQ865167.1, DQ865168.1, DQ865169.1, DQ865170.1, DQ865171.1, and DQ865172.1. Maximum parsimony (MP) and Neighbor-joining (NJ) trees (Saitou and Nei, 1987), which were based upon Tamura-

Nei (1993) distance estimates were reconstructed using MEGA v.4 (Kumar et al., 2004). Both analyses utilized all codon positions. Support for nodes was assessed using 1,000 bootstrap reiterations.

The MP and NJ analyses were repeated using an alignment of sequences varying in length from 619 to 1095 bp. LWS sequences analyzed in this analysis include the following species and acquisition numbers: Zebrafish (*Danio rerio*), AB087803.1 and AB087804.1; Japanese rice fish or medaka (*Oryzias latipes*) AB223051.1 and AB223052.1; Bluefin killifish (*Lucania goodie*) AY296740.1 and AY296741.1; Blind cave fish (*Astyanax mexicanus*) M90075.1, U12024.1, and U12025.1; Sea chub (*Girella punctata*) AB158261.1; Nile tilapia (*Oreochromis niloticus*) AF247128.1; Fugu (*Takifugu rubripes*) AY598942.1; Spotted green pufferfish (*Tetraodon nigroviridis*) AY598943.1; Turbot (*Scophthalmus maximus*) AF385826.1; Winter flounder (*Pseudopleuronectes americanus*) AY631039.1; Goldfish (*Carassius auratus*) L11867.1; Coho salmon (*Oncorhynchus kisutch*) AY214145.1; Ayu smelt LWS AYU-R (*Plecoglossus altivelis*) AB098702.1; Ayu smelt LWS Red-sensitive (*Plecoglossus altivelis*) AB107771.1; Atlantic halibut (*Hippoglossus hippoglossus*) AF316498.1; Carp (*Cyprinus carpio*) AB055656.1; human (*Homo sapiens*) NM_020061.3 and NM_000513.1; Arctic lamprey (*Lethenteron japonicum*) AB116381.1; and our new sequences from the Cumaná guppy (*Poecilia reticulata*), Picta or 'swamp guppy' (*Poecilia picta*), Parae (*Poecilia parae*), Bifurca (*Poecilia bifurca*), the Pygmy swordtail (*Xiphophorus pygmaeus*) and Tomeurus (*Tomeurus gracilis*).

Results

Hybrid or mosaic sequences

A large number of LWS-like sequences (up to 17 per guppy) were uncovered after cloning and sequencing the products of PCR reactions containing primers Fw1a and Rev5. These sequences included suspected recombinants, that is, sequences that could have been generated by the ordered concatenation of fragments of other sequences produced in the same PCR reaction. Template switching during PCR and/or mismatch repair of cloned heteroduplex molecules has been shown to generate such artefacts (Saiki et al, 1988; Oldenberg et al., 1995; Zylstra et al., 1998; Shammas et al., 2001). To test the hypothesis that PCR and cloning could generate LWS sequences not found in a genome, we used primers Fw1a and Rev5 to re-amplify DNA from a two-sequence template (i.e., two insert-bearing plasmids). Five different sequences were uncovered from the two-template PCR reaction; one copy of each of the two templates and three hybrid sequences. These two-template experiments confirmed speculation by Hoffmann et al. (2007) and Weadick and Chang (2007) that LWS opsin genes in poeciliids are susceptible to PCR and/or cloning artefacts that generate artificial hybrid sequences.

To determine the minimum number of genuine LWS opsin sequences in our dataset we first considered variation at polymorphic positions. Two sequences (e.g., two homozygous loci or two alleles at one locus) could serve as a PCR template for the generation of an enormous diversity of hybrid sequences via template switching or mismatch repair. However, among such a set of hybrid sequences there would be only two variants (substitutions or indels) at a given polymorphic site. LWS opsin sequences derived from individual fish using primers Fw1a and Rev5 included three different intron II haplotypes and a position in exon III that was polymorphic for three different nucleotides. Remarkably, this exon III variation translated into amino acid variation at position 180; the first of the five sites known to influence spectral sensitivity (see below). Gene duplication is the only explanation for the occurrence of three haplotypes in a single individual. We set out to strengthen this evidence for LWS opsin gene duplication by amplifying DNA between the genes.

Between-gene PCR and sequencing

PCR using primers designed to amplify between-gene DNA (Fw1a Comp and Rev8 Comp) produced a ~4 Kbp product in the guppy and in the three species of the guppy sister group Micropoecilia (*P. picta*, *P. parae*, and *P. bifurca*). These amplicons were cloned in three of these four species (cloning of this amplicon was unsuccessful in *P. bifurca*) and approximately 1500 bp were sequenced from each end of the clone insert. Each end contained the last intron and exon of an LWS opsin gene and approximately 790 bp beyond the stop codon. The explanation for this sequence pattern, confirmed by subsequent PCR experiments using only Rev8 Comp, was that this fragment was amplified with Rev8 Comp acting as a forward and a reverse primer and that this amplicon contained the ends of two LWS loci oriented in an inverted (tail-to-tail) fashion. The between-gene fragment did not amplify from *X. pygmaeus* or *T. gracilis*. In the guppy, additional primers were designed and the entire intergenic sequence was characterized. It was 3329 bp long, 66% A/T, and contained a short compound microsatellite; $(TGGA)_{10}(TA)_9$.

LWS opsins in the family Poeciliidae

Given the evidence for the artificial generation of opsin sequence variation during PCR or cloning, and the observation that artefacts produced by template switching and/or mismatch repair do not appear to be reproducible (Zylstra et al., 1998; Cronn et al., 2002), only haplotypes recovered from multiple independent PCR and cloning experiments were assumed to represent genuine opsin sequences. Additional primers were designed from these reliable sequences and from the sequences obtained by the between-gene PCR experiments described above (see table 1.1).

Initially, three different LWS opsin sequences were identified in the guppy (*Poecilia reticulata*). These three genes were delimited by variation at codon 180; TCT (serine), GCT (alanine), and CCT (proline), and by unique intron II and intron V mutations. Codon 180 is one of the five key positions that influence wavelength sensitivity (Yokoyama and Radlwimmer, 2001) and the variation uncovered here is reflected in the names we have given to each of the loci; LWS S180 (S for serine), LWS A180 (A for alanine), and LWS P180 (P for proline).

Seven LWS S180 sequences were obtained from six different Cumaná guppies with five of these including the start codon, all exons and introns, and part of the 3' UTR (see figure 1.1). Thirteen LWS A180 sequences were obtained from seven guppies. Only one was full-length and six included sequence from exon II to the 3' UTR. The LWS

A180 sequence appears to be a naturally occurring hybrid in the Cumaná guppy; the first five exons and four introns are most similar to LWS S180, whereas the last intron and exon are identical to the LWS P180 locus. As the two regions of this hybrid LWS A180 sequence will give a conflicting phylogenetic signal, the LWS A180 sequences were truncated in the phylogenetic analyses reported below (i.e. only the first five exons and first four introns were utilized). Seven LWS P180 sequences were obtained from seven guppies. The ForBeg primer, which includes the start codon, combined with any of the reverse primers, did not amplify the LWS P180 locus. Therefore, LWS P180 sequences spanned exon II to the 3' UTR. In addition to the proline residue at site 180, the LWS P180 locus has amino acids substitutions at two other key sites. In guppies, LWS P180 also possessed a variable-length tetranucleotide microsatellite in intron III. PCR experiments using Fw100 and a reverse primer complementary only to Weadick and Chang's (2007) variant 6 sequence (RevA) uncovered a fourth LWS opsin. While we did not amplify or sequence the first exon or intron, we show that the rest of this gene is intronless, suggesting that variant 6 is a single-exon gene, arising from a retrotransposition. It has a serine at position 180 (codon: TCG) and is named LWS S180r (S for the serine at position 180 and r for retrotransposition). Finally, our southern blot shows four bands (see figure 1.2), consistent with the PCR-based hypothesis that the Cumaná guppy has four LWS loci. These four LWS opsins have the following five key-site haplotypes: SHYTA (LWS S180 and LWS S180r), AHYTA (LWS A180), and PHFAA (LWS P180) and are thus expected to be most sensitive to three different wavelengths of light (Yokoyama and Radlwimmer, 2001).

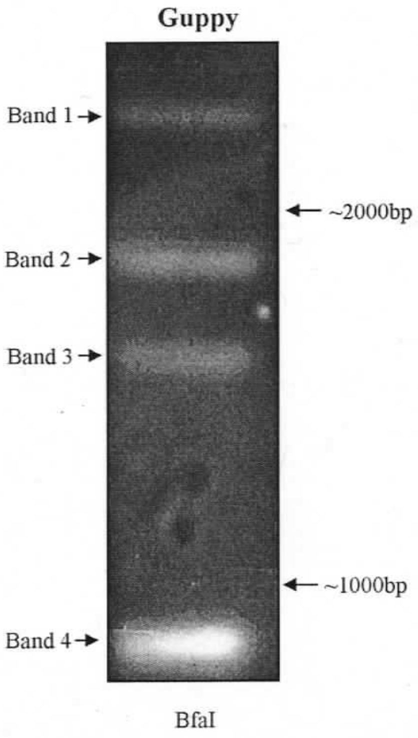


Figure 1.2. *Southern blot hybridization for determination of LWS opsin copy-number in Poecilia reticulata.* Four bands (labeled 1-4) correlate with four LWS loci from a single Cumaná guppy. BfaI (New England Biolabs®) was used as a restriction enzyme. A generic DIG-labeled probe was designed to target each LWS locus (see methods). Two size markers (given in base-pairs) are shown to the right of the blot.

Long portions of the four LWS genes found in the guppy were also amplified and sequenced from *P. picta* and *P. bifurca*. Three of these opsins, LWS S180, LWS S180r, and LWS P180, were sequenced from *P. parae*. We did not obtain the 3' end of LWS A180 from any of these three species. Therefore the mutation producing the hybrid sequence consistently recovered from Cumaná guppy cannot yet be mapped onto the poeciliid phylogeny. *Xiphophorus pygmaeus* had three LWS genes; LWS S180, LWS S180r, and LWS P180. Only one LWS sequence (LWS S180) was recovered from *Tomeurus gracilis*. All sequences have been deposited in GenBank under accession numbers EU329428 - EU329486 (see table 1.2).

Phylogenetic analysis of LWS gene duplication in Poeciliidae

Phylogenetic analysis of the 390 bp multiple sequence alignment sorted the guppy LWS opsins into three well supported clades; LWS S180r, LWS P180 and LWS S180 plus LWS A180. Weadick and Chang's (2007) variant 6 clustered with the single exon gene LWS S180r, and variant 5 clustered with the LWS P180 gene. This last result was anticipated before phylogenetic reconstruction because when translated, both sequences have a phenylalanine at position 277 and an alanine at position 285. The remaining guppy LWS genes; variants 1-4 from Weadick and Chang (2007, the three LWS opsins from Hoffman et al. (2007) (LWS_OR6-4_D09, LWS_OR6-3_E08, LWS_QUEm5_L06) and our LWS S180 and LWS A180 genes, form the third cluster. The relationships among these sequences could not be resolved. Over this 390 bp alignment, these genes are almost identical (mean percent identity = 98.6%). We suspect that the Weadick and Chang (2007) variants 1-4 and the three Hoffman et al. (2007) LWS genes include alleles at the LWS A180 and LWS S180 loci. This guppy-only LWS sequence comparison also revealed that Weadick and Chang's (2007) variant 5 is a hybrid or mosaic sequence; the first 221 bp are identical to variant 4, and the last 170 bp are identical to our LWS P180 sequence. This gene occurs in the LWS P180 clade because the region where it is identical to variant 4 is a region with few phylogenetically informative characters.

Maximum parsimony analysis of the longer alignment with the new poeciliid sequences and LWS opsins from a diversity of ray-finned fish produced a single tree (see figure 1.3). The topology of the NJ tree included all of the nodes from the MP tree that have bootstrap support >65% and many of the nodes with lower support. Unlike the MP tree in figure 1.3, the NJ analysis placed the *Tomeurus* LWS opsin as the sister sequence to a clade with the *Xiphophorus* and *Poecilia* LWS P180, LWS S180 and LWS A180 genes. This reconstruction makes more sense than the MP tree with respect to poeciliid taxonomy; morphological and molecular data indicate that *Poecilia* and *Xiphophorus* are more closely related to one another than either is to *Tomeurus*. However, we present the MP tree because the neighbor-joining tree also placed the bluefin killifish LWSA gene at the base of the LWS S180r clade and the *Xiphophorus* LWS S180 gene at the base of the LWS P180 clade. The number of gene duplication events and gene losses required to reconcile such a topology with the well-supported taxonomic relationships among these species makes these components of the NJ topology very unlikely.

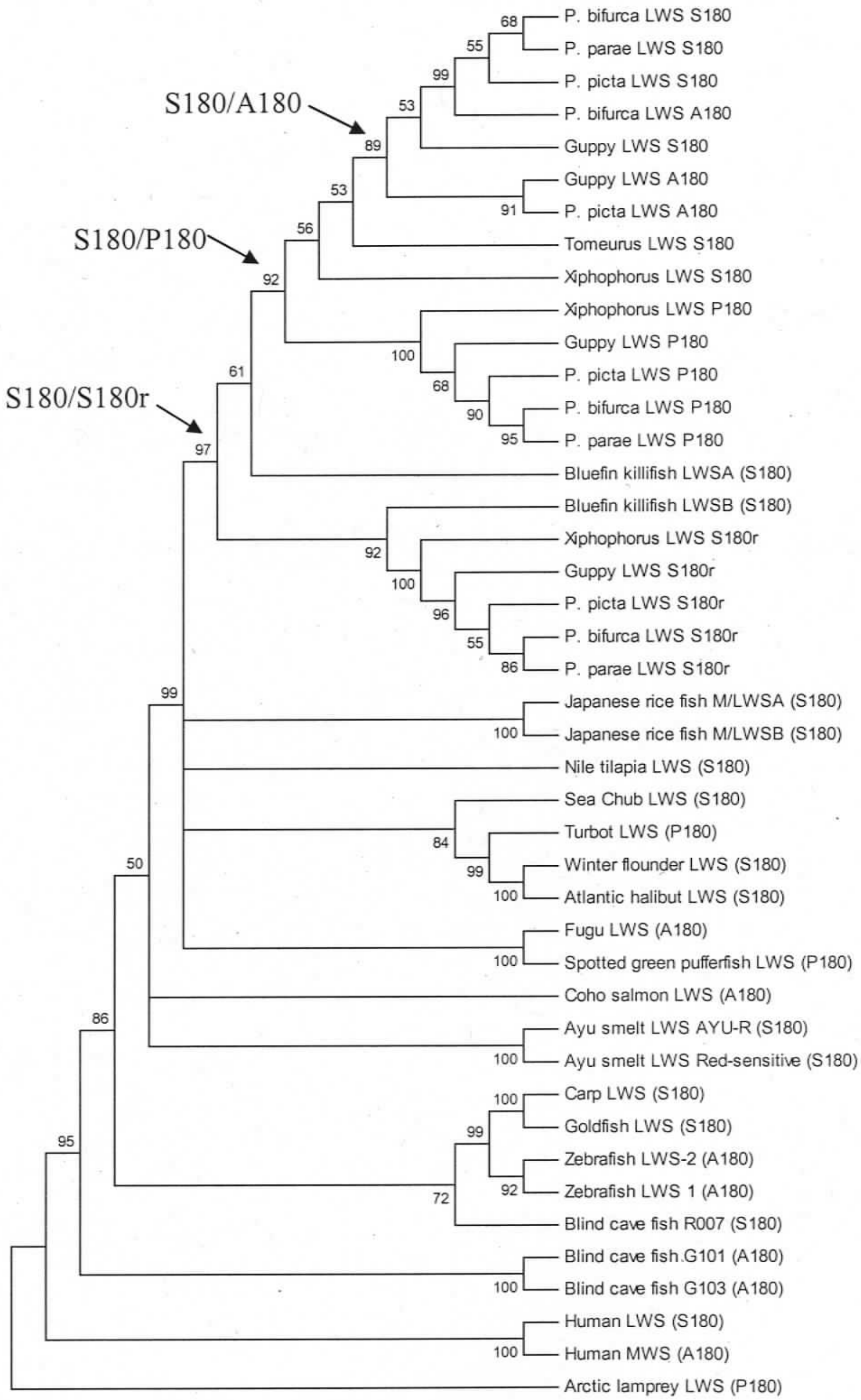

S180/A180
S180/P180
S180/S180r
P. bifurca LWS S180
P. parae LWS S180
P. picta LWS S180
P. bifurca LWS A180
Guppy LWS S180
Guppy LWS A180
P. picta LWS A180
Tomeurus LWS S180
Xiphophorus LWS S180
Xiphophorus LWS P180
Guppy LWS P180
P. picta LWS P180
P. bifurca LWS P180
P. parae LWS P180
Bluefin killifish LWSA (S180)
Bluefin killifish LWSB (S180)
Xiphophorus LWS S180r
Guppy LWS S180r
P. picta LWS S180r
P. bifurca LWS S180r
P. parae LWS S180r
Japanese rice fish M/LWSA (S180)
Japanese rice fish M/LWSB (S180)
Nile tilapia LWS (S180)
Sea Chub LWS (S180)
Turbot LWS (P180)
Winter flounder LWS (S180)
Atlantic halibut LWS (S180)
Fugu LWS (A180)
Spotted green pufferfish LWS (P180)
Coho salmon LWS (A180)
Ayu smelt LWS AYU-R (S180)
Ayu smelt LWS Red-sensitive (S180)
Carp LWS (S180)
Goldfish LWS (S180)
Zebrafish LWS-2 (A180)
Zebrafish LWS 1 (A180)
Blind cave fish R007 (S180)
Blind cave fish G101 (A180)
Blind cave fish G103 (A180)
Human LWS (S180)
Human MWS (A180)
Arctic lamprey LWS (P180)
68
55
99
53
89
91
53
56
92
100
68
90
95
61
97
92
100
96
55
86
99
100
84
99
100
50
100
86
100
100
99
92
72
95
100
100

Figure 1.3. *Phylogenetic analysis of LWS opsin genes in ray-finned fish.* A Maximum Parsimony (MP) bootstrap consensus tree of long wavelength-sensitive (LWS) opsins from representative ray-finned fish lineages. The percentage of trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown at the nodes. Nodes with less than 50% bootstrap support are collapsed. This MP tree was obtained using the Close-Neighbor-Interchange algorithm (Nei and Kumar, 2000) with search level 3 (Felsenstein, 1985; Tamura et al., 2007) in which the initial trees were obtained with the random addition of sequences (10 replicates). All codon positions were included and gaps treated as missing data. There were 1101 positions in the alignment and 520 were parsimony informative. The tree was rooted with the arctic lamprey LWS gene. LWS duplication events are marked with an asterix. Phylogenetic analyses utilized MEGA4 (Tamura et al., 2007).

The MP tree indicates that the guppy LWS opsin repertoire is a consequence of gene duplication events that occurred i) prior to the divergence of families Fundulidae and Poeciliidae, ii) in the common ancestor of *Xiphophorus* and *Poecilia*, and iii) within the genus *Poecilia.* Mechanisms of LWS opsin gene duplication include retrotransposition (producing LWS S180r) and inverted tandem duplication (producing the gene pair, LWS S180 and LWS P180). Formation of the hybrid LWS A180 locus possibly involved quasipalindrome correction (van Noort et al., 2003). These three duplication events have provided *Poecilia* species with a larger repertoire of LWS pigments than any other fish taxon.

LWS gene evolution in teleosts

Relationships among higher taxonomic groups were well-resolved in the tree reconstructed from LWS opsin sequences (see figure 1.3). There is high (>75%) bootstrap support for monophyly of Cyprinidontiformes (the bluefin killifish and all poeciliids), Pleuronectiformes (sea chub, turbot and flounder), Percomorpha, and the family Cyprinidae (goldfish, carp and zebrafish). One of the blind cavefish (*Astyanax mexicanus*) LWS genes (R007) was the sister sequence to those from goldfish, carp and zebrafish, which is not surprising as all species are in the taxon Ostariophysi. However, there were also two cavefish LWS genes at the base of the actinopterygian clade (G101 and G103). These genes might be derived from a gene produced during the fish-specific whole genome duplication event (Taylor et al., 2003). Long Branch Attraction (LBA) occurs when rapidly evolving sequences are attracted to the base of a tree (Bergsten, 2005) and is an alternative explanation for the position of the cavefish duplicates in our analysis. However, LBA is an artefact that is usually correlated with poor taxonomic sampling, which is not the case here.

Among LWS sequences available for all ray-finned fishes, there is much variation at the five sites that influence spectral sensitivity most. Three different amino acids were observed at position 180 (A, S, or P), two at position 277 (Y or F), and two at position 285 (T or A). SHYTA is believed to be the ancestral five-site haplotype for vertebrates (Yokoyama and Radlwimmer, 2001). Serine to alanine substitutions at position 180 are common, but only guppies, turbot, and the spotted green pufferfish have a proline in this position. Lamprey, though not a ray-finned fish, also has a proline at position 180. Amino acid variation at positions 197 and 308 are also known to influence spectral sensitivity, however, all ray-finned fish surveyed to date possess only H197 and A308. LWS opsins have been duplicated six times in ray-finned fishes; twice within Poeciliidae, once prior to the divergence of Fundulidae and Poeciliidae, once in medaka (*Oryzias latipes*), zebrafish (*Danio rerio*), and again in the blind cave fish (*Astyanax mexicanus*). Only in poeciliids and zebrafish (and humans), has the duplication been followed by a substitution at one or more of the key sites (see table 1.3).

Key Site Position	Guppy LWS S180	Guppy LWS A180	Guppy LWS P180	Guppy LWS S180r	Killifish LWS A	Killifish LWS B	Rice fish LWS A	Rice fish LWS B	Zebrafish LWS 1	Zebrafish LWS 2	Cave fish LWS g101	Cave fish LWS g103	Human LWS	Human MWS
180	S	A	P	S	S	S	S	S	A	A	A	A	S	A
197	H	H	H	H	H	H	H	H	H	H	H	H	H	H
277	Y	Y	F	Y	Y	Y	Y	Y	Y	F	F	F	Y	F
285	T	T	A	T	T	T	T	T	T	T	A	A	T	A
308	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Expected λ_{max} (nm)	~560	~553	~531 +/-P?	~560	~560	~560	~560	~560	~553	~546	~531	~531	~560	~531

Table 1.3. *Long wave-sensitive opsin key-site haplotypes for LWS duplicated vertebrates.* Shaded squares represent key-site amino acid substitutions from the ancestral SHYTA sequence (Yokoyama and Radlwimmer, 2001). The λ_{max} sensitivity of each LWS gene is predicted based upon the expected spectral shift caused by a key-site substitution. The functional shift of a proline substitution at amino acid position 180 is currently not known. Accession numbers are listed in the methods section.

LWS opsin gene expression

RT-PCR experiments (see table 1.2) show that all four transcripts were expressed at the same time in the eyes of adult guppies. We then used qPCR to compare transcript copy numbers. In three adult guppies, LWS A180 was expressed at a much higher level than LWS S180 and LWS S180r. LWS P180 was expressed at very low levels (see figure 1.4). The amount of cDNA in the three samples was estimated by comparing critical threshold (Ct (dR)) values between samples and standard curves prepared from plasmids containing each transcript (see table 1.4 and 1.5). Standard curve log fit values are shown in table 1.5.

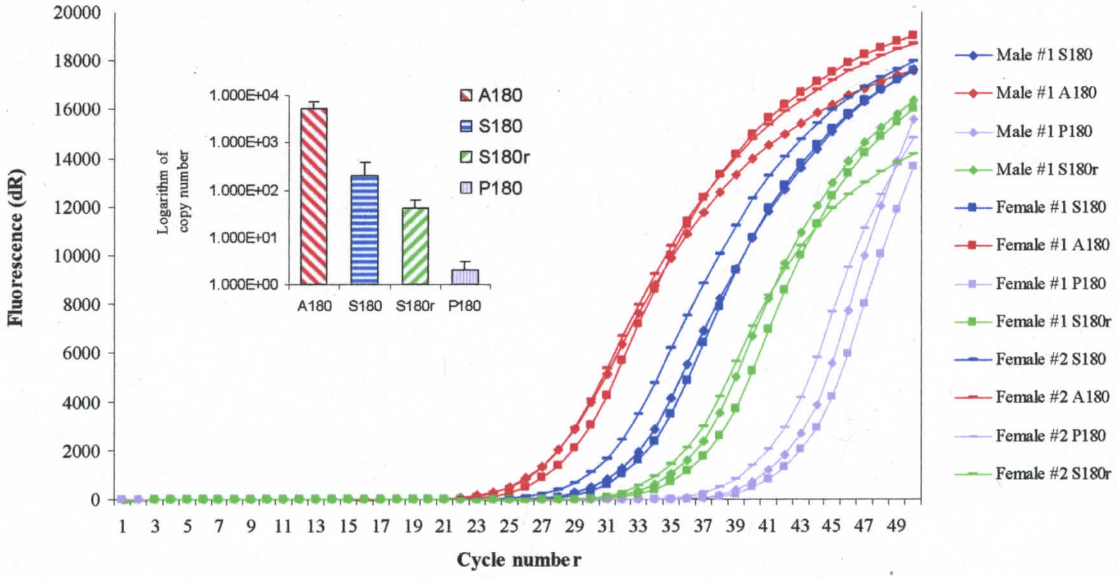


Figure 1.4. *Quantitative PCR.* Four LWS transcripts from the cDNA of three adult guppies were quantified using qPCR with the amplification plot shown. Ct (dR) values are given in table 1.4 (threshold of 370.3). A histogram showing the averaged logarithm value of original transcript copy number is shown (with standard error bars included).

Individual	Transcript	Ct (dR)	Original Copy Number
Male #1	A180	24.35	6758
Male #1	S180	29.54	121
Male #1	S180r	33.12	39
Male #1	P180	39.01	1+
Female #1	A180	25.46	3126
Female #1	S180	30.10	78
Female #1	S180r	33.90	24
Female #1	P180	39.58	1+
Female #2	A180	24.49	6131
Female #2	S180	27.93	416
Female #2	S180r	32.36	63
Female #2	P180	37.51	1+

Table 1.4. *Quantitative PCR results.* Ct (dR) values (threshold level of 370.3) are shown for each LWS opsin transcript from three adult guppies. Original copy number was calculated using Ct values and the standard curve described in table 1.5.

(Plasmid) Standard Curve	Linear Equation	Efficiency (%)	RSq
A180	Y= -3.315(LogX) + 37.05	100.3	0.999
S180	Y= -2.992(LogX) + 35.77	115.9	0.996
S180r	Y= -3.670(LogX) + 38.96	87.3	0.985
P180	Y= -3.319(LogX) + 36.68	100.1	0.999

Table 1.5. *Standard curve log fit values for quantitative PCR.* The linear equation, percentage efficiency and r-squared (RSq) values for the qPCR standard curve are shown (threshold level of 370.3). Standard plasmid copy numbers were determined by calculating the number of moles in each sample multiplied by Avogadro's number. A standard curve of Ct (dR) was used to calculate original copy number of mRNA transcripts in the diluted cDNA samples.

Discussion

Variation at key sites among LWS opsin gene duplicates in Poeciliidae

The guppy (*Poecilia reticulata*) and species in its sister group Micropoecilia possess four LWS opsin genes that we have named LWS S180, LWS S180r, LWS A180, and LWS P180. The first two genes encode proteins with the five key-site haplotype, SHYTA, and the second two encode proteins with haplotypes, AHYTA and PHFAA respectively. Three of these genes; LWS S180, LWS S180r, and LWS P180, were also amplified and sequenced from the genome of the pygmy swordtail, *Xiphophorus pygmaeus*. We found only the LWS S180 gene in *Tomerus gracilis*. The proline residue at key site 180 in LWS P180 might disrupt the transmembrane domain (Pakula and Sauer, 1989; Weitz et al., 1992) and compromise opsin protein function (Weitz et al., 1992; Gunther et al., 2006). However, several observations suggest that it is functional. First, LWS P180 is at least 70 million years old (it evolved before the divergence of *Poecilia* and *Xiphophorus*) and it has no other amino acid substitutions that are expected to disrupt function. Second, the LWS P180 locus has diverged from paralogous LWS opsins in ways that are expected to enhance colour vision; positions 277 and 285 have experienced a tyrosine to phenylalanine and threonine to alanine substitution, respectively. These are the same key-site substitutions involved in the evolution of an MWS opsin from an LWS opsin in humans. Third, this gene is transcribed, albeit at very low levels. Finally, LWS opsins from arctic lamprey, turbot, and the spotted green pufferfish also have a proline at position 180 and no other substitutions likely to disrupt protein function.

All four LWS opsins uncovered in this study are predicted to have unique roles in colour vision. With three different five key-site haplotypes, these opsins are predicted to be most sensitive to three different wavelengths of light. Also, despite encoding a gene with the same key-site haplotype as LWS S180, the LWS S180r opsin differs from all other LWS opsins at amino acid positions known to play a role in binding and activating transducin (Weadick and Chang, 2007).

Southern blot experiments in our study revealed four bands (see figure 1.2) consistent with the hypothesis that Cumaná guppies have four LWS opsin loci. Hoffman et al. (2007) produced a southern blot with only three bands and suggested that guppies have a minimum of two LWS genes. Variation in LWS opsin gene number among populations may be another trait guppies share with humans (Wolf et al., 1999; Macke and Nathans, 1997; Jagla et al., 2002;).

Two of the four LWS opsins described here (LWS S180 and LWS A180) were reported in Hoffman et al.'s (2007) study of guppies from the Quare and Oropuche Rivers in Trinidad. Although sequence data reported by Weadick and Chang (2007) did not include all five key sites, our phylogenetic analysis indicates that Weadick and Chang (2007) sequenced portions of all four loci from the Paria River guppy. The phylogenetic relationships among guppy LWS opsin paralogs reported by Weadick and Chang (2007) differ from those shown here. Both topologies were produced using maximum parsimony, but by surveying more individuals, species and obtaining longer sequences we have produced a larger set of parsimonious-informative characters. This is most evident when considering the relationship between LWS P180, variant 5, and variant 4. The large number of differences between variant 5 and variant 4 are apparent in Weadick and Chang's (2007) tree where maximum likelihood branch lengths have been superimposed on the MP topology. Nonetheless, in their MP analysis, these two sequences form a monophyletic group. However, the sister sequence relationship between variant 5 and variant 4 disappears with the addition of LWS P180 genes from other *Poecilia* species and *Xiphophorus* because many of the unique nucleotides (autopomorphies) in variant 5 become synapomorphies (shared derived traits) in the larger dataset. Also, two of the three characters that had united variant 5 and variant 4 in Weadick and Chang's (2007) MP analysis (adenines at positions 126 and 213 in their alignment), appear to be homoplasious when compared to a much larger set of LWS P180 and LWS S180 sequences. The origin of these homoplasies is intriguing and is discussed below.

Mechanisms of LWS duplication in poeciliids

The first duplication event that expanded the guppy LWS opsin repertoire produced two genes with the SYHTA haplotype; LWS S180 and LWS S180r. Until this study, the LWS S180r gene (called variant 6 by Weadick and Chang (2007)) had been recovered only from cDNA, and only from a Paria River guppy and the blue-finned killifish, *Lucania goodie*. The LWS S180r sequences obtained here are much longer than that already reported by Weadick and Chang (2007), but are still incomplete (i.e. the sequences are missing exon I and intron I). The observation however that this gene is missing introns II-V in all poeciliids surveyed strongly suggests that this is a single exon gene and likely arising from a retrotransposition. The genomic location of S180r is unknown.

In medaka (*Oryzias latipes*) and zebrafish (*Danio rerio*), duplicated LWS opsins are linked and oriented in a head-to-tail manor (Chinen et al., 2005; Matsumoto et al., 2006). Phylogenetic analysis shows that independent mutations produced these gene pairs (Matsumoto et al., 2006). Our study also characterized an LWS opsin tandem duplication event. Duplication of the ancestral LWS S180 gene (which possessed introns) early during the evolution of poeciliids, produced an LWS opsin that retained the SHYTA haplotype (LWS S180) and an LWS opsin that evolved a PHFAA haplotype (LWS P180). These two genes are linked, but in an inverted (i.e., tail-to-tail) orientation. Several models have been proposed to explain the formation of inverted duplicates. Secondary rearrangement after duplication by unequal sister chromatid exchange is one. Another is intra-chromosomal replication slippage in *trans* (Chen et al., 2005). This occurs when the DNA polymerase reverses direction using either the nascent strand (intra-molecular strand switch) or opposite strand (inter-molecular strand switch) as a template. By running backwards, a duplicate of the just-completed sequence is produced in an inverted orientation before the polymerase switches back to the correct template. The DNA downstream of LWS S180 has strings of adenines and thymines (data not shown) that might have facilitated strand switching by the polymerase during DNA replication (Hyrien et al., 1988). The inverted arrangement of LWS opsins in the *Poecilia* genome might make them even more prone to additional duplication events (Passananti et al., 1987). Therefore, variation in LWS gene number (among species or populations) would not be surprising given the inverted orientation of these duplicates. *Xiphophorus pygmaeus* also has the LWS P180 gene. Post-duplication gene transposition or the expansion of intergenic DNA are possible explanations for our failure to amplify DNA between LWS P180 and LWS S180 in this species.

The third and most recent LWS gene duplication uncovered in our study lead to the production of LWS S180 and LWS A180. In the Cumaná guppy, the first five exons and four introns of LWS A180 are most similar to LWS S180, and the last intron, exon, and 3' UTR are identical to LWS P180. These data suggest that LWS A180 is a hybrid gene and, therefore, the duplication event might have involved LWS S180 and LWS P180. Indeed, the formation of LWS A180 in these fish might have been facilitated by the inverted tandem orientation of LWS S180 and LWS P180 (van Noort et al., 2003). The two LWS A180 genes reported by Hoffman et al. (2007) (QUEm5_L06 and OR6-3_E08) do not have the same mosaic pattern, so the hybrid version of the LWS A180 gene may be restricted to guppies from Cumaná, Venezuela. As mentioned above, variant 5, reported by Weadick and Chang (2007) is also a hybrid sequence, with approximately half of the sequence identical to either the LWS S180 or LWS A180 genes and the other half identical to the LWS P180 gene. Just as is the case with the LWS A180 reported here in the Cumaná guppy, the position of variant 5 in the short 390 bp tree depends upon which fragments are used in the phylogenetic analysis (i.e., it would occur in the LWS S180/A180 monophyletic clade if only the first 220 bp were utilized). As there are many more phylogenetically informative characters in the second half of this sequence, variant

5 was placed in the LWS P180 clade when the entire sequence was used, despite being identical to either the LWS S180 or A180 sequences over the first 220 bp.

In the MP tree based on the larger alignment (see figure 1.3), the S180 and A180 genes are not partitioned into monophyletic clades. For instance, the *P. bifurca* sequence that clusters with the LWS S180 genes has an alanine at position 180. One explanation for this observation is that the *P. bifurca* LWS A180 sequence is an allele of the LWS S180 locus. A similar situation occurs in non-African humans where a common allele of the LWS opsin locus (which typically has the SHYTA haplotype) has an alanine in position 180 and thus, an AHYTA five key-site haplotype (Winderickx et al., 1993; Verrelli and Tishkoff, 2004)

The evolutionary consequences of LWS opsin duplication in guppies

In cichlids, variation in LWS opsin sequence and expression is associated with variation in water turbidity (Carelton and Kocher, 2001; Terai et al., 2006). This observation has lead to the hypothesis that species- and population-level differences in opsin gene sequence and expression patterns represent adaptations for foraging in either turbid or clear water. These differences in spectral sensitivity are now believed to drive and/or maintain divergence in male colouration via sexual selection (Maan et al., 2006). In guppies, the evolutionary story may be similar, in that it begins with natural selection promoting opsin gene diversification. White et al. (2005) found that blocking long wavelength light significantly reduced guppy foraging rate and suggested that cues in the long wavelength region of the spectrum contribute more to prey detection, or other aspects of foraging rate, than other wavelengths do. Guppies have multiple copies of colouration genes all linked to the sex chromosomes (Lindholm et al., 2004). This genetically based colour variation may be exploiting the female's enhanced sensitivity to slight variations in colour.

Consistent with the hypothesis that extra opsins lead to better colour vision are observations from humans and mice. Among human women who are heterozygous at either the LWS or MWS locus, some appear to have a pattern of X-inactivation that leads to tetrachromacy. These women see an average of 10 colours in a spectrum, whereas trichromatic men and women see only seven (Jameson and Highnote, 2001). Thus, humans express between one and four cones opsins and with each additional opsin, there is an associated improvement in wavelength discrimination. The hypothesis that an extra opsin gene can improve wavelength discrimination without any associated revisions to neuroanatomy, was also recently supported by a study of trichromatic mice. Female mice expressing an SWS opsin and two LWS opsins (a mouse LWS opsin gene and a human LWS opsin gene) performed much better in wavelength discrimination tests than did wild-type mice with only one LWS opsin (Jacobs et al., 2007). Finally, if LWS opsin gene duplication improves motion detection, as proposed by White et al. (2005) then female guppies might also be 'pre-adapted' to evaluate the well-characterized sigmoid display; a behavior that consists of the male arching its body into a S-shape and oscillating the long axis of the body both horizontally and vertically (Luyten and Liley, 1985).

This scenario we envision differs from cichlids in that where male cichlids appear to limit their displays to parcticular colours that are most obvious to the females in their population or species, male guppies are highly polymorphic even within populations. The evolutionary strategy may be to exploit female wavelength discrimination by displaying a great diversity of colour. This hypothesis could be tested using traditional mate choice experiments and sorting males by spectrophotometer-determined wavelength variation.

An alternative explanation for LWS opsin gene diversity in guppies is that it allows the retina to be spectrally tuned for different environments (e.g., Carlton et al., 2005; Whitemore and Bowmaker, 1989; Archer et al., 1995; Pointer et al., 2005; Takechi and Kawamura, 2005; Spady et al., 2005; Temple et al., 2006). For example, eels (*Anguilla anguilla*) have two rhodopsins, each tuned to slightly different wavelengths. They express a green-shifted locus as juveniles in fresh water and a paralogous blue-shifted locus when they return to the ocean and mature (Archer et al., 1995). The lamprey (*G. australis*) also adjusts its spectral sensitivity by changing opsin gene expression as it moves from a marine to riverine environment (Davies et al., 2007). However, guppies do not move within streams very far during their lifetime (Crispo et al., 2006), and thus differential use of opsin gene duplicates in different habitat is an unlikely explanation for the evolution of LWS opsin gene diversity in this taxon.

LWS expression in the guppy

Gene expression data can help test alternative hypotheses about the adaptive value of LWS opsin gene diversity. To improve wavelength discrimination, it is necessary that different opsins be expressed at the same time. All four LWS opsin gene transcripts were amplified from cDNA derived from adult eyes in our lab and by Weadick and Chang (2007). However, our qPCR experiments on three adults (1 male, 2 females) showed that most of the LWS opsin mRNA in the Cumaná guppies retina was LWS A180. Interestingly, Human SWS (blue) cone cells make up only 15% of the retina cone cell repertoire, yet play an important role in wavelength discrimination. Thus, qPCR data showing unequal expression among LWS opsin paralogs do not rule out a role for LWS opsin gene duplication and divergence in wavelength discrimination in guppies but does indicate the need for further investigation. We are currently using qPCR to examine LWS expression in a larger sample of adults and in fish at different stages of development. Finally, duplicated opsins are also sometimes expressed in different regions of the retina (Chinen et al., 2005; Takechi and Kawamura, 2005). *In situ* hybridization (ISH) experiments are also underway to test the hypothesis that different LWS opsin gene paralogs have unique expression domains within the guppy retina, as is the case in zebrafish (Takechi and Kawamura, 2005).

Chapter 2

Supplemental Research

Abstract

The first chapter of this thesis was a modified version of a manuscript prepared for publication in BMC Evolutionary Biology. Because this paper was already quite lengthy, there was much information and data excluded from the arcticle. This additional material however represents important contributions to the general question of LWS opsin genes in the family Poeciliidae, and is presented here in chapter 2. The following chapter can therefore be thought of as an 'author's notes' section to work already presented by Ward et al. (2008) and will thus delve deeper into the challenges, findings and future directions of this work. This chapter is divided into four main sections which include; a detailed discussion of PCR artefacts encountered during LWS opsin sequencing, a discussion of LWS alleles versus LWS loci within Poeciliidae, a thorough investigation of LWS paralogs in multiple populations of the guppy, and a further analysis of paralog expression within the adult retina of this species.

Section 1: PCR Sequencing and Artificial Sequences

The first section of this chapter deals with how we sequenced the LWS opsin gene (or LWS opsin paralogs) in six species of the family Poeciliidae (*Poecilia reticulata*, *Poecilia picta*, *Poecilia bifurca*, *Poecilia parae*, *Xiphophorus pygmaeus*, and *Tomeurus gracilis*). As mentioned previously in chapter 1, during the PCR and sequencing process, we encountered multiple mosaic or artificial sequences. These artefacts were suspected of being caused either by PCR mediated template switching, or mismatch repair during *Escherichia coli* transformation and cloning. Discussed in this section is the mechanism by which both of these phenomena occur and how we demonstrated that LWS sequences in the guppy are prone to either (or both) processes. Tactics to overcome template-switching and mismatch repair correction are also considered. Finally, a strategy to determine real from artificial LWS sequences in Poeciliidae is discussed.

Introduction to PCR and PCR mediated template switching

In 1983, Kary Mullis invented a ground-breaking technique which would revolutionize molecular biology for years to come. He called his innovation the polymerase chain reaction (PCR) – a technique that had the ability to exponentially amplify a targeted DNA sequence. In 1988, Saiki et al. drastically improved this procedure with the introduction of taq; a thermal-stable polymerase isolated from *Thermus aquaticus*. This new polymerase enhanced the specificity, yield, sensitivity and length of PCR amplicons and as a result, is widely used today. It was during this time, however, that potentially serious problems with the PCR process were identified. Namely, the possibility of mosaic or 'shuffled' artefact sequences formed from partially homologous, yet non-identical templates during the amplification process (Saiki et al., 1988; Odelberg et al., 1995; Judo et al., 1998; Cronn et al., 2002). Thus, PCR-mediated template switching (or also known as 'jumping PCR') had the potential to produce a wide range of artificial sequences. Although identified early, the influence of these mosaic artefacts have received minor attention over the years and may actually be less appreciated now than in 1988. Because of the extensive use of PCR today in a multitude of fields (i.e. forensics, ecology, anthropology and medicine), and the enormous sequence databases already compiled; the notion that such an important molecular technique could have serious flaws in its processivity is a difficult concept to accept. Nonetheless, this inherent problem in the PCR process should not be ignored as it is now believed that many public databases are likely corrupt with faulty sequences as a result of PCR template switching (Hugenholtz and Huber, 2003). For instance, current estimates suggest that at least 1 in 20 16S rRNA sequences held in public repositories (such as NCBI or GenBank) are corrupt due in part to the numerous submissions of artificial mosaic sequences (Hugenholtz and Huber, 2003; Ashelford et al., 2005).

The mechanism of PCR template switching

The exact etiology and mechanism of PCR-mediated template switching is not yet fully understood, but the key requirement appears to be the inclusion of at least two partially homologous templates in one PCR reaction (Odelberg et al., 1995; Cronn et al., 2002). For instance, most genomic templates contain paralogous genes (i.e. multi-gene families) and upon targeting these duplicates for amplification, a high frequency of chimeric formation is often observed (Saiki et al., 1988). Indeed, sequence variation at a single heterozygous locus is likely to be sufficient for production of *in vitro* mosaic amplicons (Shuldiner et al., 1989; Bradley and Hillis, 1997). A model of PCR artificial recombination proposes that a proportion of incomplete extended primers anneal to heterologous templates in subsequent rounds of PCR (Shuldiner et al., 1989; Meyerhans et al., 1990; Odelberg et al., 1995; Judo et al., 1998). Incomplete extended primers may be caused by short extension times, premature polymerase dissociation or polymerase blockage due to secondary structures located within the parent strand (Shuldiner et al., 1989; Bradley and Hillis, 1997). A resulting template switch is thought to occur by one of four mechanisms. First, a partial displacement of the nascent strand(s) can result in a switch between nascent strands themselves (i.e. two extending strands meet) or to a preexisting paralog elsewhere in the template (i.e. a duplicated locus) (Odelberg et al., 1995). Alternatively, complete displacement of a nascent strand can result in a switch between other displaced nascent strands or again to a preexisting paralogous template (Odelberg et al., 1995). All four scenarios would result in a PCR amplified chimeric sequence but, may also provide an important insight into the mechanism by which an *in vivo* mosaic locus could be created via gene duplication and divergence.

Introduction to mismatch repair as an alternative process for the creation of artificial sequences

Strand specific mismatch repair (MMR) represents an important evolutionary adaptation by both prokaryotic and eukaryotic cells to correct mistakes during DNA biosynthesis and conesquently avoid detrimental mutations (Modrich, 1997; Kunkel and Erie, 2005; Iyer et al., 2006). In *E. coli*, the mechanism of MMR is well characterized and initiated when MutL-MutS binds to mismatched DNA (Nelson and Cox, 2005). Correction of the mismatched base requires identification of the nascent strand thus ensuring that the original parent sequence is used as the template (Kunkel and Erie, 2005). This is accomplished in part by Dam methylase; an important enzyme that catalyzes the methylation of adenine residues (N^6 position) in nearby 5'GATC sequences (up to ~1 Kb away) (Nelson and Cox, 2005). Newly synthesized DNA is subject to methylation after a transient delay, thus facilitating a brief hemi-methylated state in the mismatched heteroduplex (Iyer et al., 2006). MutH targets hemi-methylated 5'GATC sequences and identifies the nascent strand for correction (which temporarily lacks adenine methylation) (Iyer et al., 2006). DNA on both sides of the mismatch is threaded through the MutL-MutS complex (forming a DNA loop) until MutL encounters the bound MutH protein (Nelson and Cox, 2005). MutH then cleaves the unmethylated strand, at which time, a complex of DNA helicase II and several other exonucleases (varies dependant on which side the cleavage site is on in relation to the mismatch) degrade the nascent strand from that point towards the mismatch (Nelson and Cox, 2005). The resulting gap and mismatch is repaired by DNA polymerase III and the nick is sealed by DNA ligase (Nelson and Cox, 2005).

As proposed by L'Abbé et al. (1992), artificial mosaic sequences may not be due only to PCR mediated template switching, but could arise because of mismatch repair during the *E. coli* transformation and cloning process. When paralogous genes or alleles are PCR amplified simultaneously, a heterogeneous mixture of products is generated (Longeri et al., 2002). During the transformation and cloning process, formation of heteroduplex molecules can occur which may be subject to the *E. coli* mismatch repair pathway (L'Abbé et al., 1992; Volkov et al., 1999; Longeri et al., 2002; Thompson et al., 2002). Thompson et al. (2002) suggests that the diversity of artefact sequences generated as a result of heteroduplex formation and MMR correction will increase exponentially with the number of variable nucleotides and/or sequence variants present in the original PCR reaction.

The observation of sequence artefacts when amplifying LWS paralogs from the guppy genome

When we first began characterizing the LWS opsin gene in guppies, we uncovered a large number of LWS-like sequences (up to 17 per individual) while using conserved primers. Many of these sequences were suspected of being either PCR template-switched or MMR artefacts due to strong chimeric signals from the introns and/or strings of variable nucleotides. To test the hypothesis that PCR and MMR during cloning could generate LWS sequences not found in the guppy genome, Allison

Churcher, used primers Fw1a and Rev5 (see table 1.1) to re-amplify DNA from a two-sequence template. As shown in figure 2.1, 'Allison12' and 'Allison8' (both LWS sequences) were used as plasmid-inserted templates for the PCR amplification of five distinct sequences; three of which showed clear signals of either template switching or heteroduplex MMR correction. For instance, one of the recombinant sequences, S3mix11, possessed four distinct switches between the two templates (positions 100, 383, 425, and 837) and two locations where MMR correction may also have occurred (positions 928 and 950). This two-template experiment confirmed speculation by Hoffmann et al. (2007) and Weadick and Chang (2007) that LWS opsin genes in poeciliids are susceptible to PCR and/or MMR cloning artefacts which generate mosaic sequences.

```
                        10        20        30        40        50        60        70        80        90       100       110
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Allison12  1    TCTTATCAGTCTTCACCAACGGCCTCGTTTTGG-GGCCACAGCAAAGTTCAAGAAACTTCGTCATCCTCTCAACTGGATCTTGGTCAACCTTGCCATTGCCGATCTTGGA 109
S3mix12    1    TCTTATCAGTCTTCACCAACGGCCTCGTTTTGG-GGCCACAGCAAAGTTCAAGAAACTTCGTCATCCTCTCAACTGGATCTTGGTCAACCTTGCCATTGCCGATCTTGGA 109
S3mix11    1    TCTTATCAGTCTTCACCAACGGCCTCGTTTTGG-GGCCACAGCAAAGTTCAAGAAACTTCGTCATCCTCTCAACTGGATCTTGGTCAACCTTGCCATTGCTGATCTTGGA 109
S3mix14    1    TCTTATCAGTCTTCACCAACGGCCTCGTTTTGG-GGCCACAGCAAAGTTCAAGAAACTTCGTCATCCTCTCAACTGGATCTTGGTCAACCTTGCCATTGCCGATCTTGGA 109
S3mix13    1    TCTTATCAGTCTTCACCAACGGCCTCGTCTTGGTGGCCACAGCAAAGTTCAAGAAACTTCGTCATCCTCTCAACTGGATCTTGGTCAACCTTGCCATTGCTGATCTTGGA 110
S3mix15    1    TCTTATCAGTCTTCACCAACGGCCTCGTCTTGGTGGCCACAGCAAAGTTCAAGAAACTTCGTCATCCTCTCAACTGGATCTTGGTCAACCTTGCCATTGCTGATCTTGGA 110
Allison8   1    TCTTATCAGTCTTCACCAACGGCCTCGTCTTGGTGGCCACAGCAAAGTTCAAGAAACTTCGTCATCCTCTCAACTGGATCTTGGTCAACCTTGCCATTGCTGATCTTGGA 110

                       120       130       140       150       160       170       180       190       200       210       220
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Allison12  110  GAGACAGTCTTTGCCAGTACCATCAGTGTGTGCAACCAGGTCTTTGGATATTTCATTCTGGGACACCCAATGTGTGTCTTTGAAGGCTATGTTGTCTCAATTTGTGGTAA 219
S3mix12    110  GAGACAGTCTTTGCCAGTACCATCAGTGTGTGCA-CCAGGTCTTTGGATATTTCATTCTGGGACACCCAATGTGTGTCTTTGAAGGCTATGTTGTCTCAATTTGTGGTAA 218
S3mix11    110  GAGACAGTCTTTGCCAGCACCATCAGTGTGTGCAACCAGTTCTTTGGATATTTCATTCTGGGACACCCAATGTGTGTCTTTGAAGGCTATGTTGTCTCAACTTGTGGTAA 219
S3mix14    110  GAGACAGTCTTTGCCAGTACCATCAGTGTGTGCAACCAGGTCTTTGGATATTTCATTCTGGGACACCCAATGTGTGTCTTTGAAGGCTATGTTGTCTCAATTTGTGGTAA 219
S3mix13    111  GAGACAGTCTTTGCCAGCACCATCAGTGTGTGCAACCAGTTCTTTGGATATTTCATTCTGGGACACCCAATGTGTGTCTTTGAAGGCTATGTTGTCTCAACTTGTGGTAA 220
S3mix15    111  GAGACAGTCTTTGCCAGCACCATCAGTGTGTGCAACCAGTTCTTTGGATATTTCATTCTGGGACACCCAATGTGTGTCTTTGAAGGCTATGTTGTCTCAACTTGTGGTAA 220
Allison8   111  GAGACAGTCTTTGCCAGCACCATCAGTGTGTGCAACCAGTTCTTTGGATATTTCATTCTGGGACACCCAATGTGTGTCTTTGAAGGCTATGTTGTCTCAACTTGTGGTAA 220

                       230       240       250       260       270       280       290       300       310       320       330
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Allison12  220  GTTGTCAC------AGTGGTAAGTGAAGTACCAGACCAGTAAATAAATTTTAATTAAACCTAAAGAA---TTGTGTTGT---------------CATCTGTGCATCTATA 305
S3mix12    219  GTTGTCAC------AGTGGTAAGTGAAGTACCAGACCAGTAAATAAATTTTAATTAAACCTAAAGAA---TTGTGTTGT---------------CATCTGTGCATCTATA 304
S3mix11    220  GTTGCCACAAGTTAAGTGGTAATTTATGTTCCAGACCTGTAAATTAATTTTAATCAAACCTAAAGATGAATCATGTTTGTAGTGTGTTAAATTGCATGTGTGCATCTATA 329
S3mix14    220  GTTGTCACAAGTTAAGTGGTAATTTATGTTCCAGACCTGTAAATTAATTTTAATCAAACCTAAAGATGAATCATGTTTGTAGTGTGTTAAATTGCATGTGTGCATCTATA 329
S3mix13    221  GTTGCCACAAGTTAAGTGGTAATTTATGTTCCAGACCTGTAAATTAATTTTAATCAAACCTAAAGATGAATCATGTTTGTAGTGTGTTAAATTGCATGTGTGCATCTATA 330
S3mix15    221  GTTGCCACAAGTTAAGTGGTAATTTATGTTCCAGACCTGTAAATTAATTTTAATCAAACCTAAAGATGAATCATGTTTGTAGTGTGTTAAATTGCATGTGTGCATCTATA 330
Allison8   221  GTTGCCACAAGTTAAGTGGTAATTTATGTTCCAGACCTGTAAATTAATTTTAATCAAACCTAAAGATGAATCATGTTTGTAGTGTGTTAAATTGCATGTGTGCATCTATA 330

                       340       350       360       370       380       390       400       410       420       430       440
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Allison12  306  TTTCTGTACATAGGAATTGCTGGGCTTTGGTCCTTGACCATCATCTCTTGGGAAAGATGGATAGTTGTGTGCAAACCCTTTGGAAATGTCAAGTTTGATTCCAAGTGGGC 415
S3mix12    305  TTTCTGTACATAGGAATTGCTGGGCTTTGGTCCTTGACCATCATCTCTTGGGAAAGATGGATAGTTGTGTGCAAACCCTTTGGAAATGTCAAGTTTGATTCCAAGTGGGC 414
S3mix11    330  ATCTTT---ACAGGTATTGCTGCTCTGTGGTCCCTGACTATTATCTCTTGGGAAAGATGGATAGTTGTGTGCAAACCCTTTGGAAATGTCAAGTTCGATGCCAAGTGGGC 436
S3mix14    330  ATCTTT---ACAGGTATTGCTGCTCTGTGGTCCCTGACTATTATCTCTTGGGAGAGATGGATAGTTGTGTGCAAACCTTTTGGAAATGTCAAGTTCGATGCCAAGTGGGC 436
S3mix13    331  ATCTTT---ACAGGTATTGCTGCTCTGTGGTCCCTGACTATTATCTCTTGGGAGAGATGGATAGTTGTGTGCAAACCTTTTGGAAATGTCAAGTTCGATGCCAAGTGGGC 437
S3mix15    331  ATCTTT---ACAGGTATTGCTGCTCTGTGGTCCCTGACTATTATCTCTTGGGAGAGATGGATAGTTGTGTGCAAACCTTTTGGAAATGTCAAGTTCGATGCCAAGTGGGC 437
Allison8   331  ATCTTT---ACAGGTATTGCTGCTCTGTGGTCCCTGACTATTATCTCTTGGGAGAGATGGATAGTTGTGTGCAAACCTTTTGGAAATGTCAAGTTCGATGCCAAGTGGGC 437

                       450       460       470       480       490       500       510       520       530       540       550
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Allison12  416  CACAGCTGGAATACTTTTCTCCTGGGTCTGGCCTGCAGTGTGGTGCGCTCCTCCCATCTTTGGATGGAGCAGGTAATCAATCAATCAATAAATCTATCAATCAGTTTTAT 525
S3mix12    415  CACAGCTGGAATACTTTTCTCCTGGGTCTGGCCTGCAGTGTGGTGCGCTCCTCCCATCTTTGGATGGAGCAGGTAATCAATCAATCAATAAATCTATCAATCAGTTTTAT 524
S3mix11    437  CACAGCTGGAATAGTTTTCTCCTGGGTCTGGGCTGCAGTGTGGTGCGCTCCTCCCATCTTTGGATGGAGCAGGTAATC-------------------------------- 514
S3mix14    437  CACAGCTGGAATAGTTTTCTCCTGGGTCTGGGCTGCAGTGTGGTGCGCTCCTCCCATCTTTGGATGGAGCAGGTAATC-------------------------------- 514
S3mix13    438  CACAGCTGGAATAGTTTTCTCCTGGGTCTGGCCTGCAGTGTGGTGCGCTCCTCCCATCTTTGGATGGAGCAGGTAATCAATCAATCAATAAATCTATCAATCAGTTTTAT 547
S3mix15    438  CACAGCTGGAATAGTTTTCTCCTGGGTCTGGGCTGCAGTGTGGTGCGCTCCTCCCATCTTTGGATGGAGCAGGTAATC-------------------------------- 515
Allison8   438  CACAGCTGGAATAGTTTTCTCCTGGGTCTGGGCTGCAGTGTGGTGCGCTCCTCCCATCTTTGGATGGAGCAGGTAATC-------------------------------- 515

                       560       570       580       590       600       610       620       630       640       650       660
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Allison12  526  TTATAAAGCACCCTTCATACTAAAAGCAGCTCAAAGTGCTGTACAGATCAATAAAAACAAACAATTATCCCCACCCAACTCCCCCGGGTACCACATGGTAATATACTGGA 635
S3mix12    525  TTATAAAGCACCCTTCATACTAAAAGCAGCTCAAAGTGCTGTACAGATCAATAAAAACAAACAATTATCCCCACCCAACTCCCCCGGGTACCACATGGTAATATACTGGA 634
S3mix11    514  -------------------------------------------------------------------------------------------------------------- 514
S3mix14    514  -------------------------------------------------------------------------------------------------------------- 514
S3mix13    548  TTATAAAGCACCCTTCATACTAAAAGCAGCTCAAAGTGCTGTACAGATCAATAAAAACAAACAATTATCCCCACCCAACTCCCCCGGGTACCACATGGTAATATACTGGA 657
S3mix15    515  -------------------------------------------------------------------------------------------------------------- 515
Allison8   515  -------------------------------------------------------------------------------------------------------------- 515

                       670       680       690       700       710       720       730       740       750       760       770
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Allison12  636  TCTTCTGGTTATGAAGAGTTTATATACTCAATAATCCGCAGAGCAAGTTTTGTTTAATGCAAAGCAAATGCTGAAATATTTCCTTGTATCAGGTATTGGCCTCATGGACT 745
S3mix12    635  TCTTCTGGTTATGAAGAGTTTATATACTCAATAATCCGCAGAGCAAGTTTTGTTTAATGCAAAGCAAATGCTGAAATATTTCCTTGTATCAGGTATTGGCCTCATGGACT 744
S3mix11    514  ----TAG-TTATATGGTTTTTATATCCTCGATAATCAGCAGAGCAAGAACTGATTG----------------------TTTCCTTGTATCAGGTATTGGCCTCATGGACT 597
S3mix14    514  ----TAG-TTATATGGTTTTTATATCCTCGATAATCAGCAGAGCAAGAACTGATTG----------------------TTTCCTTGTATCAGGTATTGGCCTCATGGACT 597
S3mix13    658  TCTTCTGGTTATGAAGAGTTTATATACTCAATAATCCGCAGAGCAAGTTTTGTTTAATGCAAAGCAAATGCTGAAATATTTCCTTGTATCAGGTATTGGCCTCATGGACT 767
S3mix15    515  ----TAG-TTATATGGTTTTTATATCCTCGATAATCAGCAGAGCAAGAACTGATTG----------------------TTTCCTTGTATCAGGTATTGGCCTCATGGACT 598
Allison8   515  ----TAG-TTATATGGTTTTTATATCCTCGATAATCAGCAGAGCAAGAACTGATTG----------------------TTTCCTTGTATCAGGTATTGGCCTCATGGACT 598

                       780       790       800       810       820       830       840       850       860       870       880
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Allison12  746  GAAAACATC-TGCGGACCTGATGTGTTCAGTGGAAGTGAAGACCCTGGAGTCCAGTCCTACATGATGGTCCTCATAATTACATGCTGCTTCATTCCTCTGGCTATCATCA 854
S3mix12    745  GAAAACATC-TGCGGACCTGATGTGTTCAGTGGAAGTGAAGACCCTGGAGTCCAGTCCTACATGATGGTCCTCATAATTACATGCTGCTTCATTCCTCTGGCTATCATCA 853
S3mix11    598  GAAAACATCCTGCGGACCTGATGTGTTCAGTGGAAGCGATGACCCAGGGGTCCTGTCCTACATGATGGTCCTCATAATTACATGCTGCTTCATTCCTCTGGCTATCATCA 707
S3mix14    598  GAAAACATC-TGCGGACCTGATGTGTTCAGTGGAAGCGATGACCCAGGGGTCCTGTCCTACATGATTGTCCTCATGATTACATGCTGCATCATTCCTCTGGCTATCATCA 706
S3mix13    768  GAAAACATC-TGCGGACCTGATGTGTTCAGTGGAAGTGAAGACCCTGGAGTCCAGTCCTACATGATGGTCCTCATAATTACATGCTGCATCATTCCTCTGGCTATCATCA 876
S3mix15    599  GAAAACATCCTGCGGACCTGATGTGTTCAGTGGAAGCGATGACCCAGGGGTCCTGTCCTACATGATTGTCCTCATGATTACATGCTGCATCATTCCTCTGGCTATCATCA 708
Allison8   599  GAAAACATCCTGCGGACCTGATGTGTTCAGTGGAAGCGATGACCCAGGGGTCCTGTCCTACATGATTGTCCTCATGATTACATGCTGCATCATTCCTCTGGCTATCATCA 708

                       890       900       910       920       930       940       950       960       970       980       990
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Allison12  855  TCTTGTGCTACCTGGCTGTGTGGTTGGCCATCCGTGCTGTAAGTTTT-ATCCCTCATTGCTTTTTTTTC-TTCAGATTTATTGTAGCACTAAAGAAATCTTGATCAAATT 962
S3mix12    854  TCTTGTGCTACCTGGCTGTGTGGTTGGCCATCCGTGCTGTAAGTTTT-ATCCCTCATTGCTTTTTTTTC-TTCAGATTTATTGTAGCACTAAAGAAATCTTGATCAAATT 961
S3mix11    708  TCTTGTGCTACCTGGCTGTGTGGTTGGCCATCCGTGCTGTAAGTTTTTATCCCTCATTGCTTTTTTTTCCTTCAGATTTATTGTAGCACTAAAGAAATCTTGATCAAATT 817
S3mix14    707  TCTTGTGCTACCTGGCTGTGTGGTTGGCCATCCGTGCTGTAAGTTTTTATCCCTCATTGCTTTTTTTTCCTTCAGATTTATCGTAGCACTAAAGAAAGTTTGTTAAAATT 816
S3mix13    877  TCTTGTGCTACCTGGCTGTGTGGTTGGCCATCCGTGCTGTAAGTTTT-ATCCCTCATTGCTTTTTTTTCCTTCAGATTTATCGTAGCACTAAAGAAAGTTTGTTAAAATT 985
S3mix15    709  TCTTGTGCTACCTGGCTGTGTGGTTGGCCATCCGTGCTGTAAGTTTTTATCCCTCATTGCTTTTTTTTCCTTCAGATTTATCGTAGCACTAAAGAAAGTTTGTTAAAATT 818
Allison8   709  TCTTGTGCTACCTGGCTGTGTGGTTGGCCATCCGTGCTGTAAGTTTTTATCCCTCATTGCTTTTTTTTCCTTCAGATTTATCGTAGCACTAAAGAAAGTTTGTTAAAATT 818

                      1000      1010      1020      1030      1040      1050      1060      1070      1080      1090
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Allison12  963  TGACCTCATGTTTTGTTGACTACAGGTTGCTATGCAGCAGCTAGATAGTGAATCAACTCAGAAGGCTGAGAGAGAAGTGTCCAGGATGGTTGTAGTCATG 1062
S3mix12    962  TGACCTCATGTTTTGTTGACTACAGGTTGCTATGCAGCAGCTAGATAGTGAATCAACTCAGAAGGCTGAGAGAGAAGTGTCCAGGATGGTTGTAGTCATG 1061
S3mix11    818  TGACCTCATGTTTTGTTGACTACAGGTTGCTATGCAGCAGCTAGATAGTGAATCAACTCAGAAGGCTGAGAGAGAAGTGTCCAGGATGGTTGTAGTCATG 917
S3mix14    817  TGACTTCATGTTTTGTTGGCTACAGGTTGCTATGCAGCAGAAGGAATCTGAGTCGACCCAGAAGGCTGAGAGAGAAGTGTCCAGGATGGTTGTAGTCATG 916
S3mix13    986  TGACTTCATGTTTTGTTGGCTACAGGTTGCTATGCAGCAGAAGGAATCTGAGTCGACCCAGAAGGCTGAGAGAGAAGTGTCCAGGATGGTTGTAGTCATG 1085
S3mix15    819  TGACTTCATGTTTTGTTGGCTACAGGTTGCTATGCAGCAGAAGGAATCTGAGTCGACCCAGAAGGCTGAGAGAGAAGTGTCCAGGATGGTTGTAGTCATG 918
Allison8   819  TGACTTCATGTTTTGTTGGCTACAGGTTGCTATGCAGCAGAAGGAATCTGAGTCGACCCAGAAGGCTGAGAGAGAAGTGTCCAGGATGGTTGTAGTCATG 918
```

Figure 2.1. *Two-template PCR experiment.* Shown is a sequence alignment (BioEdit v.7.0.5.3.) demonstrating evidence for the production of artificial LWS opsin sequences during a two-template PCR reaction. It is unclear if the mosaic sequences (S3mix11, S3mix14, S3mix13) were generated during the polymerase chain reaction (via template switching) or during transformation and cloning of competent cells (via mismatch repair). Allison 12 (blue) and Allison 8 (red) were used as the original templates and were amplified from Cumaná guppy DNA using primers Fw1a and Rev5 (see table 1.1).

Strategies to overcome the formation of artefacts caused by PCR-mediated template switching or mismatch repair during cloning

Many studies have offered practical solutions to PCR template switching and MMR problems, but to date, no clear protocol to avoid either phenomena completely exists. There are however, some useful modifications that can be done both during PCR and the cloning process, which may help to minimize the presence of artificial sequences. For instance, Polz et al. (1998) suggests that to help avoid PCR bias and template switching, degenerate positions in primer target sites should be avoided when using universal primers; as subtle differences in melting curves may influence mismatch hybridization during annealing. Furthermore, to increase reproducibility, PCR should be carried out using initial high template concentrations in conjunction with replicate reactions, thus minimizing PCR drift (Wagner et al., 1994; Polz et al., 1998; Acinas et al., 2005). To minimize mosaic sequences and limit polymerase errors, it has been suggested that the smallest possible number of PCR amplification cycles should also be used (Suzuki and Giovannoni, 1996; Polz et al., 1998; Acinas et al., 2005). Even the choice of polymerase can make a drastic difference in both fidelity and processivity during the PCR reaction (Qiu et al., 2001). Finally, long extension times, high annealing temperatures and a low ramp rate between the denaturation and annealing steps, should all be avoided when trying to limit template switching during PCR (Ishii and Fukui, 2001; Kurata et al., 2004; Acinas et al., 2005).

Strategies to decrease the frequency of chimeric sequences caused by MMR during cloning are less developed, although a key requirement appears to be the reduction of heteroduplexes formed within the PCR product (Thompson et al., 2002). Heteroduplex molecules are more likely to form during PCR when the targeted DNA has multiple, highly similar sequences within the same template (Thompson et al., 2002). Targeting only one paralogous locus using specific primers (such as those complementary to UTR) is more likely to facilitate homoduplex formation within the final PCR product, which will be less prone to MMR correction during transformation and cloning.

Characterization of Poeciliidae LWS opsin paralogs

Initially, the Poeciliidae LWS sequences we identified as being artificial chimeras resembled a process of PCR drift. That is, independent of the etiology of the artefacts (i.e. template switching or MMR), these sequences showed random signals of mosaicism. Therefore, as discussed in chapter 1, the best approach we felt to properly characterize 'real' LWS genes in Poeciliidae was to amplify and sequence repeatedly from the same genetic locus using multiple individuals, PCR products and clones. Added to this strategy was the need to reduce the frequency of mosaic artefacts in order to build a proper dataset. We accomplished this by using many of the PCR template switching and MMR strategies discussed above. For instance, by using high PCR template concentrations (~100ng per reaction), short extension times, and a high fidelity polymerase (iProof DNA polymerase from BioRad®), production of mosaic artefacts appeared to be reduced. Specific primers complementary to unique LWS loci also helped to lower the frequency of artefact production; perhaps by decreasing the number of heteroduplex molecules formed in the reaction and thus limiting MMR correction during cloning. Unique reverse primers for LWS S180 and LWS P180 were based on their respective 3'UTRs (see chapter 1). A unique reverse primer for LWS S180r was based on the variant 6 sequence recently published by Weadick and Chang (2007) (see chapter 1). Using these primers on their respective loci seemed drastically decrease the frequency of PCR artefacts within our final dataset.

Using multiple alignments, we thoroughly inspected each of our LWS opsin sequences for clear signals of artificial mosaicism. As shown in figure 2.1., artefacts can be identified by observing small or large strings of nucleotide switches between real sequences, and in turn, form non-reproducible hybrids. After considering individual variation, SNPs, and the likelihood for polymerase/sequencing errors, we identified 71 real LWS sequences from our Poeciliidae samples (see table 1.2). A Jukes and Cantor (1969) tree is shown below to demonstrate the distance between each of these LWS sequences when grouped into locus-specific clades (see figure 2.2). Once we were confident that the sequences we possessed in our final dataset were real and could not be attributed to PCR drift, a pairwise and complete deletion similarity matrix was used to compute the p-distance of each locus group. These results are summarized below in table 2.1, which provides further evidence that our filtered LWS sequences are in fact real. Other than the two LWS S180 sequences from *Poecilia bifurca* (which are suspected alleles at this locus), each locus-specific group with more than two sequences show a mean >99% identity. Because we sequenced the S180r locus only once from five species of Poeciliidae, these sequences were analyzed together, yet still show a >97% identity. Locus sequences that we obtained only once were compared to their closest relatives and again, signals of potential artefacts were carefully evaluated. Because of the high degree of identity among each locus group, it is unlikely that any of these sequences exhibit the random mosaic patterns we had seen in our previously discarded sequences.

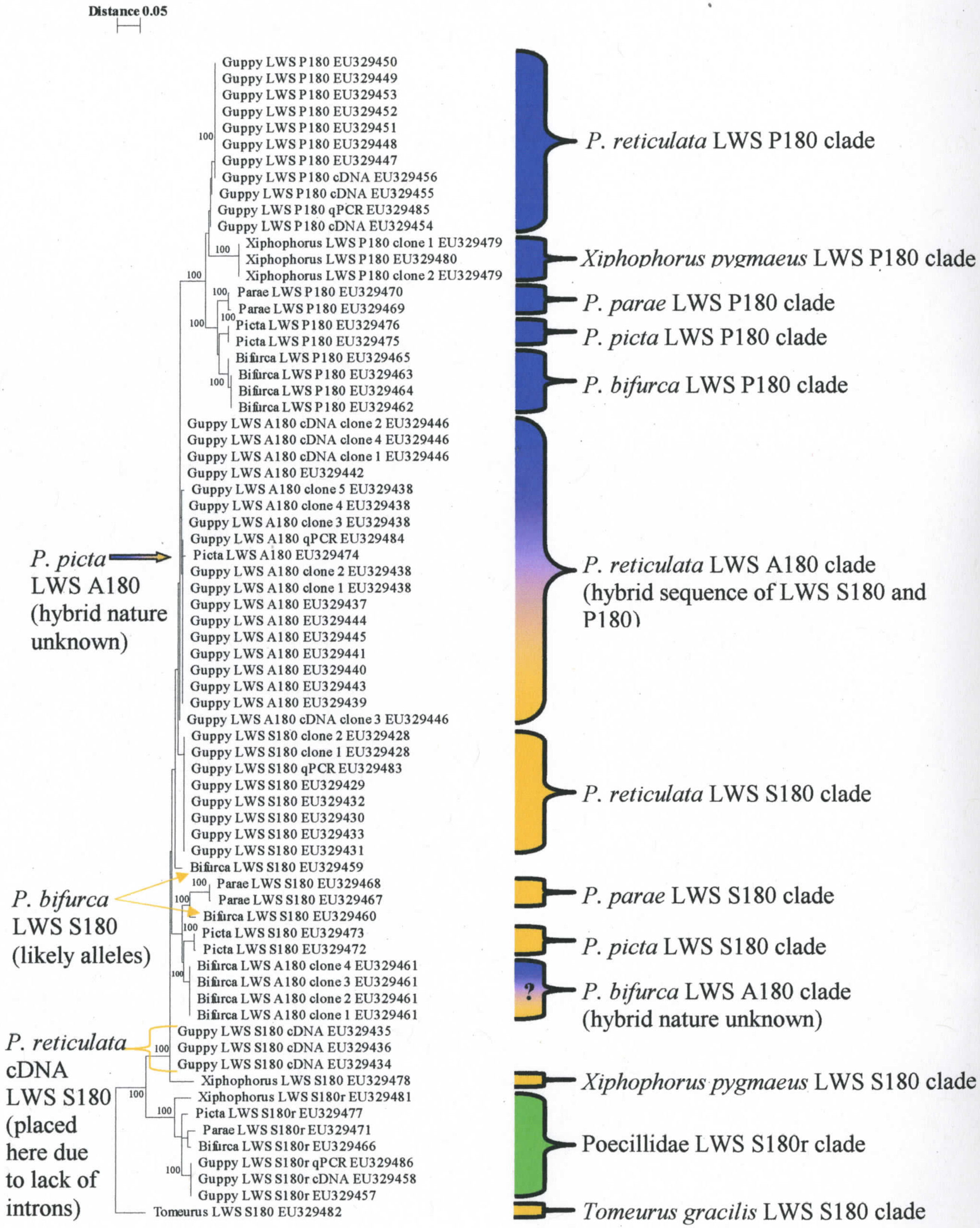

Distance 0.05
Guppy LWS P180 EU329450
Guppy LWS P180 EU329449
Guppy LWS P180 EU329453
Guppy LWS P180 EU329452
Guppy LWS P180 EU329451
Guppy LWS P180 EU329448
Guppy LWS P180 EU329447
Guppy LWS P180 cDNA EU329456
Guppy LWS P180 cDNA EU329455
Guppy LWS P180 qPCR EU329485
Guppy LWS P180 cDNA EU329454
Xiphophorus LWS P180 clone 1 EU329479
Xiphophorus LWS P180 EU329480
Xiphophorus LWS P180 clone 2 EU329479
Parae LWS P180 EU329470
Parae LWS P180 EU329469
Picta LWS P180 EU329476
Picta LWS P180 EU329475
Bifurca LWS P180 EU329465
Bifurca LWS P180 EU329463
Bifurca LWS P180 EU329464
Bifurca LWS P180 EU329462
Guppy LWS A180 cDNA clone 2 EU329446
Guppy LWS A180 cDNA clone 4 EU329446
Guppy LWS A180 cDNA clone 1 EU329446
Guppy LWS A180 EU329442
Guppy LWS A180 clone 5 EU329438
Guppy LWS A180 clone 4 EU329438
Guppy LWS A180 clone 3 EU329438
Guppy LWS A180 qPCR EU329484
Picta LWS A180 EU329474
Guppy LWS A180 clone 2 EU329438
Guppy LWS A180 clone 1 EU329438
Guppy LWS A180 EU329437
Guppy LWS A180 EU329444
Guppy LWS A180 EU329445
Guppy LWS A180 EU329441
Guppy LWS A180 EU329440
Guppy LWS A180 EU329443
Guppy LWS A180 EU329439
Guppy LWS A180 cDNA clone 3 EU329446
Guppy LWS S180 clone 2 EU329428
Guppy LWS S180 clone 1 EU329428
Guppy LWS S180 qPCR EU329483
Guppy LWS S180 EU329429
Guppy LWS S180 EU329432
Guppy LWS S180 EU329430
Guppy LWS S180 EU329433
Guppy LWS S180 EU329431
Bifurca LWS S180 EU329459
Parae LWS S180 EU329468
Parae LWS S180 EU329467
Bifurca LWS S180 EU329460
Picta LWS S180 EU329473
Picta LWS S180 EU329472
Bifurca LWS A180 clone 4 EU329461
Bifurca LWS A180 clone 3 EU329461
Bifurca LWS A180 clone 2 EU329461
Bifurca LWS A180 clone 1 EU329461
Guppy LWS S180 cDNA EU329435
Guppy LWS S180 cDNA EU329436
Guppy LWS S180 cDNA EU329434
Xiphophorus LWS S180 EU329478
Xiphophorus LWS S180r EU329481
Picta LWS S180r EU329477
Parae LWS S180r EU329471
Bifurca LWS S180r EU329466
Guppy LWS S180r qPCR EU329486
Guppy LWS S180r cDNA EU329458
Guppy LWS S180r EU329457
Tomeurus LWS S180 EU329482
100
100
100
100
100
100
100
100
100
100
100
100
100
100
100
P. picta LWS A180 (hybrid nature unknown)
P. bifurca LWS S180 (likely alleles)
P. reticulata cDNA LWS S180 (placed here due to lack of introns)
P. reticulata LWS P180 clade
Xiphophorus pygmaeus LWS P180 clade
P. parae LWS P180 clade
P. picta LWS P180 clade
P. bifurca LWS P180 clade
P. reticulata LWS A180 clade (hybrid sequence of LWS S180 and P180)
P. reticulata LWS S180 clade
P. parae LWS S180 clade
P. picta LWS S180 clade
?
P. bifurca LWS A180 clade (hybrid nature unknown)
Xiphophorus pygmaeus LWS S180 clade
Poecillidae LWS S180r clade
Tomeurus gracilis LWS S180 clade

Figure 2.2. *Jukes and Cantor (1969) neighbor-joining (NJ) distance tree of Poeciliidae LWS opsin genes.* The tree was constructed using TreeCon v.1.3.b. (Van de Peer and Wachter, 1994) and a BioEdit v.7.0.5.3. (Hall, 1999) CLUSTALw alignment of all Poeciliidae LWS opsin sequences submitted to GenBank from this study (see figure 1.1 and table 1.2). All nucleotide positions were used and support for nodes was evaluated by 1000 bootstrap reiterations. Clades are colour coded: blue for the LWS P180 clades, yellow for the LWS S180 clades, green for the LWS S180r clade, and a mixture of blue and yellow for the LWS A180 clade showing the hybrid nature of this locus in the Cumaná guppy.

Species name and LWS Locus	Number of Sequences Acquired	Overall Mean of p-Distance (Pairwise Deletion)	Standard Deviation of p-Distance (Pairwise Deletion)	Overall Mean of p-Distance (Complete Deletion)	Standard Deviation of p-Distance (Complete Deletion)
P. reticulata LWS S180	10	0.000770	0.000426	0.000578	0.000555
P. reticulata LWS A180	17	0.005135	0.001407	0.003807	0.001507
P. reticulata LWS P180	10	0.002900	0.000874	0.001536	0.000879
P. bifurca LWS S180	2	0.044444	0.006996	0.044444	0.006858
P. bifurca LWS A180	4	0.000552	0.000568	0.000552	0.000524
P. bifurca LWS P180	4	0.003389	0.000978	0.000709	0.000707
P. parae LWS S180	2	0.001524	0.001054	0.001524	0.001062
P. parae LWS P180	2	0.008328	0.002206	0.008328	0.002163
P. picta LWS S180	2	0.004405	0.002239	0.004405	0.001824
P. picta LWS A180	1	-	-	-	-
P. picta LWS P180	2	0.000000	0.000000	0.000000	0.000000
Xip. pygmaeus LWS S180	1	-	-	-	-
Xip. pygmaeus LWS P180	3	0.000893	0.000659	0.000893	0.000636
Tom. gracilis LWS S180	1	-	-	-	-
Family Poeciliidae LWS S180r	6	0.030016	0.004327	0.029533	0.004034

Table 2.1. *Overall mean p-distance of LWS locus specific groups in the family Poeciliidae.* Similarity matrixes were produced using MEGA v.4 (Kumar et al., 2004) and overall mean p-distance values were calculated using both pairwise and complete deletion. All nucleotide positions were used (including a bootstrap of 1000) and standard deviation of p-distance values are shown.

Section 2: Copy-Number Determination of Guppy LWS Opsin Paralogs

The second section of this chapter deals with how we determined that the Cumaná guppy (and likely its close relatives in Poeciliidae) has four LWS opsin loci. Recent publications by Hoffman et al. (2007) and Weadick and Chang (2007) suggest that these fish may have three and six LWS opsin loci, respectively; results inconsistent with our southern blot, genomic and expressional analyses.

Southern-based determination of LWS opsin copy-number in the guppy genome

The southern blot shown in chapter 1 (see figure 1.2) possesses four distinct bands; consistent with the hypothesis that the Cumaná guppy has four LWS opsin loci. On this blot, a second individual was analyzed, but the quality of the bands was not clear enough to be included in the chapter 1 manuscript. Nonetheless, four bands were observed in this second fish, which is shown in figure 2.3. To confirm that the copy-number of LWS paralogs suggested by the first southern blot was not unique to these two fish, or a product of experimental error, a second southern blot was repeated on an additional two Cumaná guppies using the same reaction conditions (see figure 2.3). Again, four bands were observed in each individual.

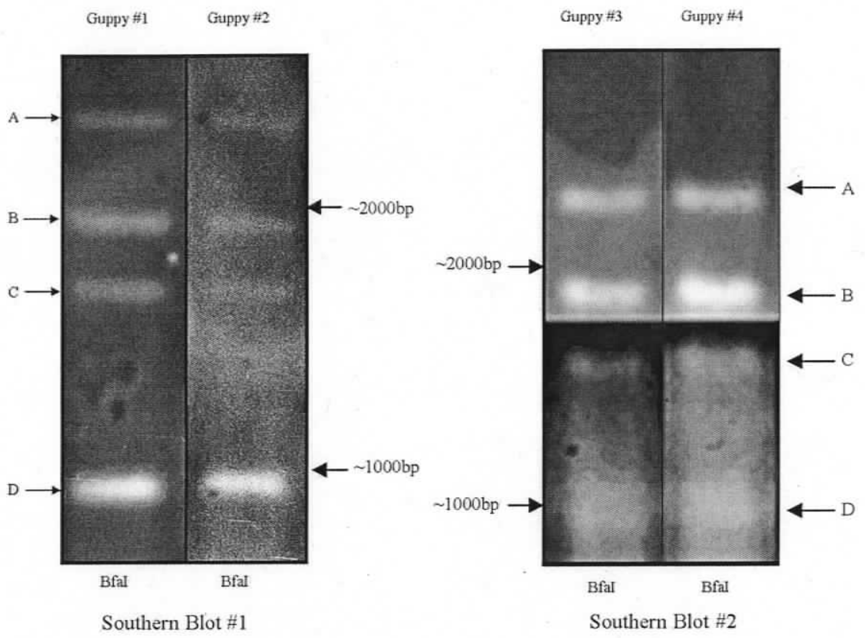


Figure 2.3. *LWS opsin southern blots of four Cumaná guppies.* Experimental conditions for both blots are described in the methods section of chapter 1. The blot shown for guppy #1 is the same as that presented in figure 1.2. Photoshop CS3 was used to invert the colour contrast in the bottom half of southern blot #2. This allowed for a better resolution of bands C and D. The restriction enzyme BfaI was used to digest the guppy genomic DNA.

Of interest, Hoffman et al. (2007) have claimed that their southern blot data supports the hypothesis of three LWS opsin loci in their population samples of Tranquille (Tr), Cumaná (Cu) and Quare River (Qu) guppies. In figure 2(a) of their paper; an EcoRI restriction digest shows two bands from a male and female individual, and a BamHI restriction digest shows three. What is peculiar is that in figure 2(b) and 2(c) of their paper (see figure 2.4 below), restriction digests using HindIII and PvuII, respectively, and hybridized with an LWS probe, show more than three bands in their LWS opsin southern blots. For instance, figure 2(c) of their paper (parcticularly in the Quare River and Cumaná populations) shows what appears to be a banding pattern of four; with pairs of bands suggesting alleles and/or restriction length polymorphisms at each LWS locus. Therefore, the results of the Hoffman et al. (2007) southern blots in conjunction with our own southern blot data, help to support the notion that guppies, across multiple populations, may posses four LWS loci, and not three.

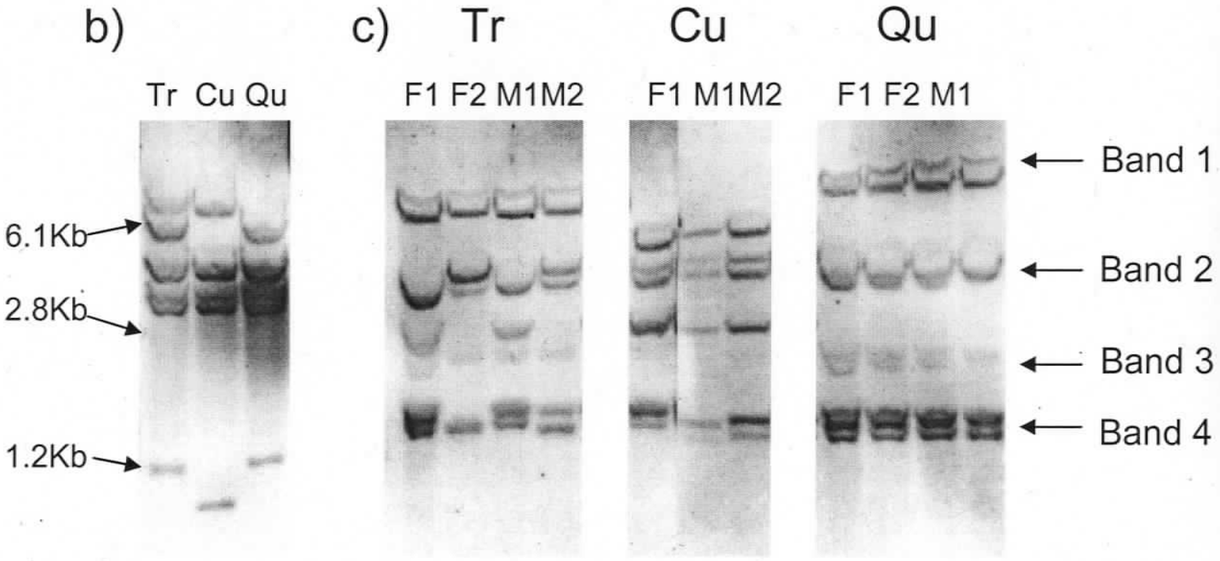


Figure 2.4. *Hoffman et al. (2007) southern blots (figure 2b and 2c) for LWS opsin copy-number in three populations of the guppy; Tranquille (Tr), Cumaná (Cu) and Quare River (Qu).* Southern blot (b) represents pooled DNA from multiple individuals (in each population) using the restriction enzyme HindIII. Southern blots (c) represent DNA collected from individual females and males (separated by population) using the restriction enzyme PvuII. Examples of band pairs (1-4) are labeled to right of the Quare River population in the later southern blot.

PCR-based determination of LWS opsin copy-number in the guppy genome

Southern blot data from our study and Hoffman et al. (2007) have provided strong evidence for four LWS opsin loci in the guppy. However, it is the genomic characterization of these paralogs which has provided the best evidence for this copy-number of LWS duplicates. For instance, from our initial PCR experiments using only Rev8 Comp (see results of chapter 1); we have determined that the LWS S180 and LWS P180 paralogs are indeed, separate loci. This is because both duplicates are tandemly linked in an inverted orientation, and in turn, separated by ~3.3 Kbp of intergenic DNA. PCR experiments which amplified the LWS S180r paralog have also shown that this duplicate must exist at a separate genomic locus. This is because the LWS S180r duplicate is intronless, and thus could not be at either the LWS S180 or LWS P180 loci. At this point, three LWS opsin loci were known to exist within the Cumaná guppy's genome, but it remained unclear whether the LWS A180 sequence represented a fourth locus or an allele at one of the already identified loci.

As discussed in chapter 1, the LWS A180 sequence was best amplified by using the LWS2 IntRev primer in conjunction with one of our universal LWS opsin forward primers. In the Cumaná guppy, this combination simultaneously amplifies two LWS duplicates: LWS A180 and LWS P180. This phenomenon is observed during gel electrophoresis, where two bands correlating to each sequence and separated by ~200bp, are consistently observed. The rationale for why this occurs during the PCR process is that LWS2 IntRev binds to the 3'UTR of both sequences because LWS A180 is a hybrid of LWS S180 (5' end) and LWS P180 (3' end).

Initially, co-amplification of LWS P180 and LWS A180 using LWS2 IntRev appeared to show an allelic pattern. That is, a banding pattern of either or both duplicates in a given PCR reaction suggested that LWS A180 could be an allele of LWS P180. However, upon re-amplification of any Cumaná guppy showing apparent homozygosity, both bands (correlating to LWS P180 and LWS A180) could eventually be amplified. Furthermore, by using a family pedigree (see figure 2.5) consisting of multiple F1 and F2 progeny, we were consistently unable to recover any individual that possessed only one of these two bands after re-amplification. Because of the large sample size of this family, and the observation that none of our in-lab bred Cumaná guppies could be shown to posses only one of these sequences, we were able to conclude that the LWS A180 duplicate must represent a separate genomic locus. We attributed the observation of apparent alleles during LWS P180 and LWS A180 co-amplification to early-cycle PCR bias. That is, during the preliminary cycles of PCR, either locus may have been detected first by the reverse primer, thus facilitating a larger initial copy number of the now-biased paralog. As demonstrated in figure 3 of Small et al. (2004), when using universal primers to co-amplify paralogous genes, subtle changes in PCR conditions (such as minute changes in $MgCl_2$ or annealing temperatures) can drastically alter the amplification of the different paralogs. Slight variations in reagent volumes between PCR reactions (possibly due to human error) or subtle changes to PCR extension times, may have facilitated the observed early-cycle PCR bias for either LWS locus (see figure 2.5).

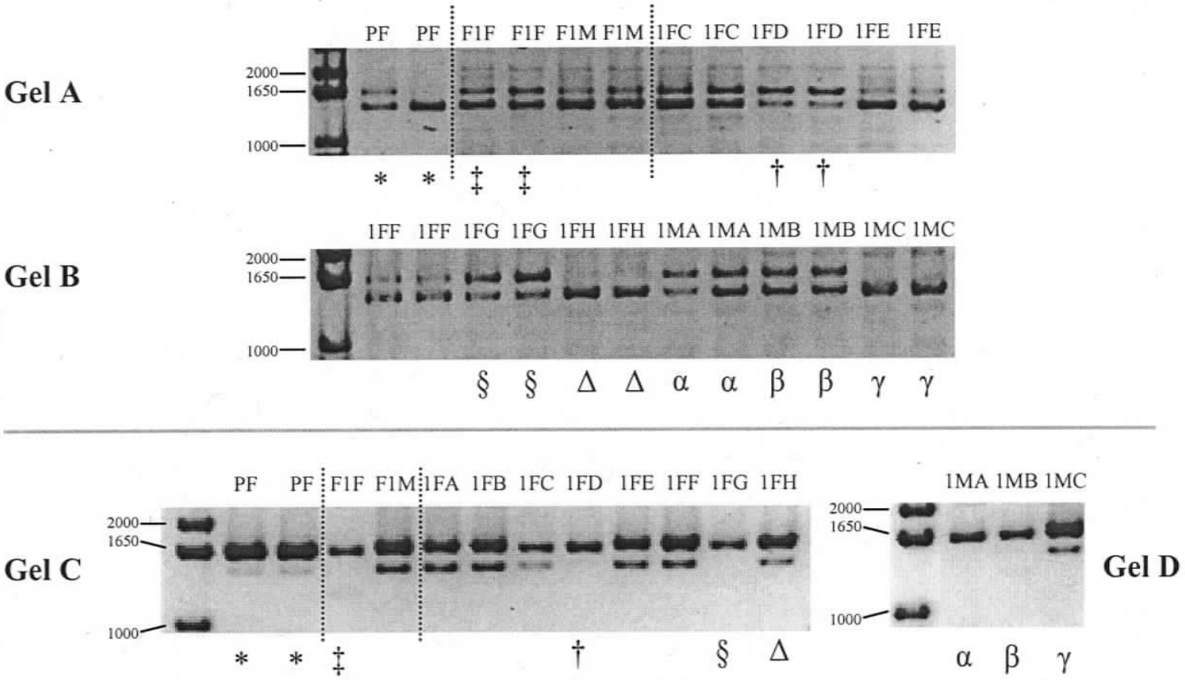


Figure 2.5. *Results of gel electrophoresis for a Cumaná guppy family pedigree of LWS A180 and LWS P180.* This experiment was performed to determine if these sequences are allelic or separate loci. Fw100 and LWS2 IntRev were used to simultaneously amplify LWS A180 (~1400 bp band) and LWS P180 (~1650 bp band) from an inbred family. PF represents the mother of F1F (female) and F1M (male), which were crossed to produce progeny 1FA to 1FH (females) and 1MA to 1MC (males). Each amplification had approximately the same PCR reagent conditions: 0.5U iProof™ DNA polymerase (BioRad), 5X iProof™ HF Buffer, 10mM dNTP mix, ~100ng template DNA, 0.5μM of For100 and LWS2 IntRev primers and dH_2O. However, PCR reaction times differed for gels A and B ([94°C(10s) + 59°C(20s) + 72°C(45s)]x35) compared to gels C and D ([94°C(10s) + 57°C(25s) + 72°C(65s)]x35). Symbols correlate to individuals which have showed apparent homozygous patterns. An Invitrogen™ 1 Kb Plus DNA ladder was used in each gel and sizes are reported in base pairs.

Because of the southern blot results demonstrated in this study, in conjunction with those shown by Hoffman et al. (2007), four LWS opsin loci are expected to exist in the guppy genome. Sequencing and pedigree analyses have now revealed that the duplicates LWS S180, LWS A180, LWS P180 and LWS S180r are indeed, separate loci. We have therefore characterized the LWS paralogs within the guppy genome, and in turn, provided important insight into the LWS duplicate repertoire of these fish.

Section 3: Analysis of LWS Opsin Sequences Published by Hoffman et al. (2007) and Weadick and Chang (2007).

As discussed in chapter 1, gene duplication and divergence has provided *Poecilia* with four distinct LWS opsin genes; a larger repertoire than any other vertebrate. Furthermore, our detailed phylogenetic analyses have shown that three of these LWS paralogs (LWS S180r, LWS P180 and LWS S180) occur in at least 239 species of the subfamily Poeciliinae. In the guppy, two previous studies (Hoffman et al., 2007; Weadick and Chang, 2007) have claimed to characterize the LWS opsin repertoire in these fish, yet their conclusions differ significantly from one another and to our own results. In the third section of this chapter, we will carefully analyze the LWS opsin sequences published by Hoffman et al. (2007) and Weadick and Chang (2007) in the hopes of clarifying any ambiguities and/or overlap to our own results.

Results from Hoffman et al. (2007) placed in the context of data presented by Ward et al. (2008)

Hoffman et al. (2007) have submitted three guppy LWS opsin cds sequences to the NCBI database, and suggest that each represents a separate genetic locus. The OR6-4_D09 sequence from the Oropuche River guppy is very similar to our Cumaná guppy LWS S180 paralog; with a p-distance of 0.0150. Variation between the OR6-4_D09 and Cumaná guppy LWS S180 sequence is likely a result of population-level genetic differences between these fish. Indeed, genetic variation at multiple loci has been shown to increase with geographic distance (Handley et al., 2007); likely explaining the apparent nucleotide substitutions observed at this locus between both studies. The other two Hoffman et al. (2007) sequences; QUEm5_L06 and OR6-3_E08, may be allelic, or represent polymorphisms at the same genetic locus. These sequences were acquired from two separate guppy populations (i.e., from the Quare and Oropuche Rivers, respectively) therefore the few subsititions observed between them likely reflect population-level divergence (see figure 2.8b). This hypothesis is supported by the fact that the 5'UTR and 3'UTR of both sequences are nearly identical, with a p-distance of 0.0172 and 0.0350, respectively. Furthermore, in the coding region of these sequences, only nine base-pair differences out of a total of 1071 nucleotides were observed and in turn, translate for only three amino acid substitutions. It appears that these sequences may represent alleles or polymorphisms at the LWS A180 locus, as both contain an alanine at amino acid 180; one of the five key-sites known to influence maximal wavelength sensitivity. Interestingly, the QUEm5_L06 and OR6-3_E08 sequences do not show the hybrid signal observed at the LWS A180 locus of the Cumaná guppy. Thus, the gene conversion event leading to a mosaic LWS A180 locus may be unique to the Cumaná population; adding to the genetic variation that has already been observed in the guppy species. Finally, using an LWS opsin alignment based on available coding sequences between our study and Hoffman et al. (2007), it is observed that the later study did not amplify the LWS P180 or LWS S180r loci (see figure 2.6) found in other Poeciliidae.

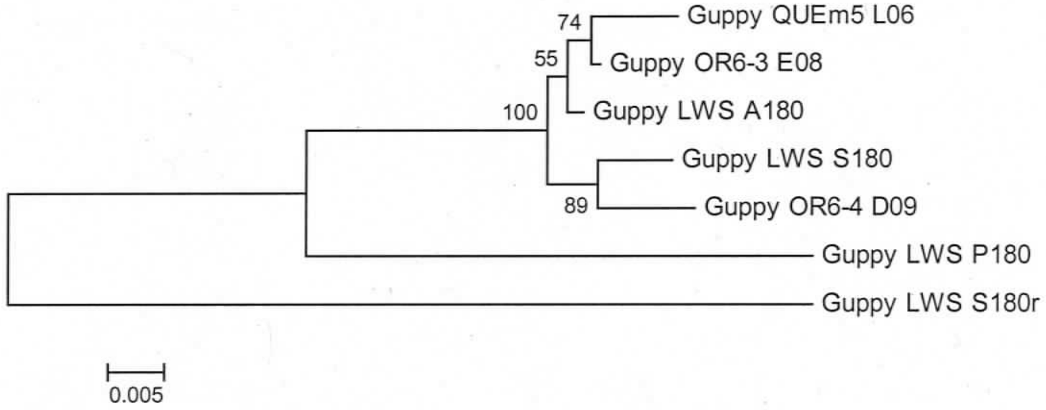


Figure 2.6. *Distance tree of LWS opsin paralogs published by our study and Hoffman et al. (2007).* The evolutionary history was inferred using MEGA4 (Tamura et al., 2007), the Neighbor-Joining / Tamura-Nei method (Saitou and Nei, 1987) and available coding sequence. The bootstrap consensus tree was inferred from 1000 replicates (Felsenstein, 1985). Codon positions included were 1st+2nd+3rd. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons with a total of 1071 positions in the final dataset. The 3’ end of the LWS A180 sequence was removed as the hybrid nature of this locus will give a false phylogenetic signal.

Results from Weadick and Chang (2007) placed in the context of data presented by Ward et al. (2008)

Six LWS opsin sequences (called variants 1 - 6) from the cDNA of a single Paria River guppy were obtained by Weadick and Chang (2007). As discussed in chapter 1, these sequences were short (~390 bp) but provided information on four of the five key-site amino acids known to influence maximal wavelength sensitivity (i.e., positions 197, 277, 285 and 308). Because of the limited size of these LWS variants, it is difficult to conduct proper phylogenetic analysis between the three studies (i.e., Hoffman et al., 2007; Weadick and Chang, 2007; Ward et al., 2008). This is parcticularly evident when considering that variants 1 – 6 have only 13 parsimony informative characters. Furthermore, many of the autapomorphic nucleotides present in these sequences become synapomorphic when aligned to a more complete LWS opsin dataset presented in this study and that published by Hoffman et al. (2007). Nonetheless, as discussed in chapter 1, a phylogenetic analysis of guppy LWS opsin sequences published in all three studies, yet limited in size to variants 1 – 6, suggest that Weadick and Chang (2007) have sequenced portions of all four loci uncovered in this study. Indeed, variant 6 and variant 5 have high sequence identity to the LWS S180r and LWS P180 loci, respectively; both found in the family Poeciliidae. Variants 1 – 4 however, have a small p-distance within its group (see figure 2.8a), and thus form a well supported monophyletic clade with the Hoffman et al. (2007) sequences and the LWS A180 and LWS S180 paralogs presented here. It is therefore likely that Weadick and Chang's (2007) variants 1-4 are alleles at these two loci; albeit without longer sequences (i.e., including key-site amino acid position 180), intronic data or UTR, it is impossible to resolve these variants/alleles into either an LWS S180 or LWS A180 clade.

Unbeknownst to the authors, Weadick and Chang (2007) have sequenced what appears to be a hybrid locus. Although it was sequenced from only one clone, and a suspicion of template switching is therefore reasonable; variant 5 is identical to variant 4 at its 5' end, and to the Cumaná guppy LWS P180 paralog at its 3' end (see chapter 1 for details). For instance, variant 5 codes for a phenylalanine and an alanine at key-site positions, 277 and 285, respectively, yet would be expected to posses either an alanine or a serine at key-site position 180. Therefore, if variant 5 represents' a true LWS sequence, then the Paria River guppy may posses a unique five-key site haplotype not yet observed in the family Poeciliidae (i.e. an A/SHFAA haplotype instead of the PHFAA haplotype found in the Cumaná guppy and other Poeciliidae). Much like the gene conversion event leading to the hybrid LWS A180 paralog in the Cumaná guppy, this mosaic variant 5 could provide further evidence of an unique population-level adaptation by the guppy to genetically diverge their LWS duplicates to provide new spectral haplotypes. It is well documented that female guppies of different populations have different preference for male colouration, perhaps contributing to the divergence of morphological patterns among populations (Houde, 1988; Houde and Hankes, 1997). If different populations of the guppy have changed their spectral genotype through LWS opsin gene conversion, this may help to explain the observed trend of variation in female guppy colour detection (Archer et al., 1987). Such a notion has important implications for the Fisher runaway process, where variation in perception could result in disparity of female choice criterion

(Archer et al., 1987); an important concept when considering the variety of spectral environments that different populations of the guppy exist in.

Consensus of data presented by Hoffman et al. (2007), Weadick and Chang (2007) and Ward et al. (2008)

To summarize the work presented in this study, Hoffman et al. (2007), and Weadick and Chang (2007), figure 2.7 illustrates the expected five key-site haplotypes of the characterized LWS opsin repertoire of four guppy populations. Of interest, it appears that the genotype of the LWS paralogs differ among these populations and may therefore allow for variation in spectral sensitivity. As discusses above, this notion has important evolutionary implications for sexual selection within these guppy populations and will be discussed in greater detail throughout the final section of this chapter. Finally, because phylogenetic analysis between these three studies is so difficult (as described in chapter 1), figure 2.8 demonstrates the respective pairwise p-distances of the LWS paralogs presented by Ward et al. (2008), Hoffman et al. (2007) and Weadick and Chang (2007). Pairwise p-distance scores can be useful in grouping respective sequences into expected loci or allelic clades, and are therefore presented here to clarify any sequence ambiguities presented by either Weadick and Chang (2007) and/or Hoffman et al. (2007).

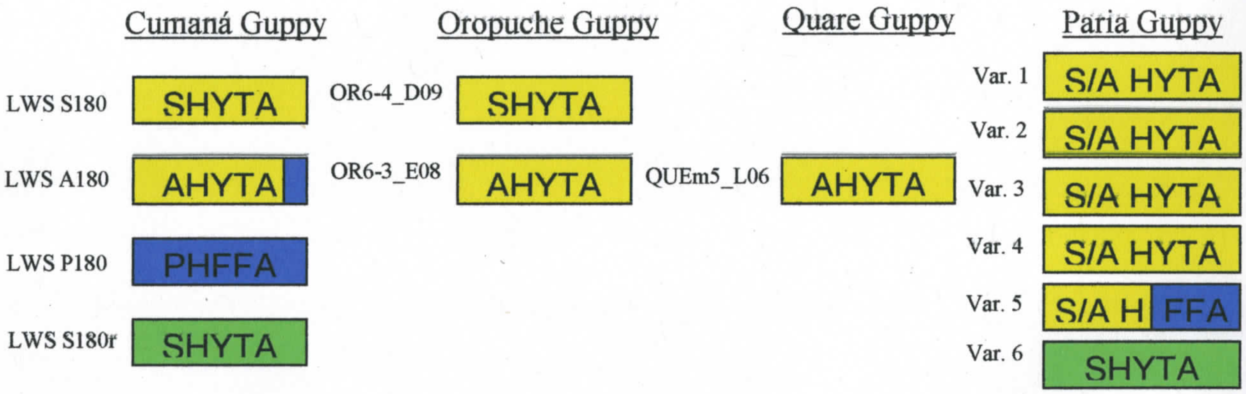


Figure 2.7. *Summary of the LWS opsin repertoire currently characterized by Ward et al. (2007), Hoffman et al. (2007), and Weadick and Chang (2007).* It is expected that the guppy genome across all populations has four LWS loci (see chapter 2, section 2 above). The LWS S180 and LWS A180 paralogs are highlighted in yellow, LWS P180 in blue and LWS S180r in green. Hybrid colours are shown to qualitatively describe mosaicism and the characterized or expected key-site amino acids are listed within their respective boxes.

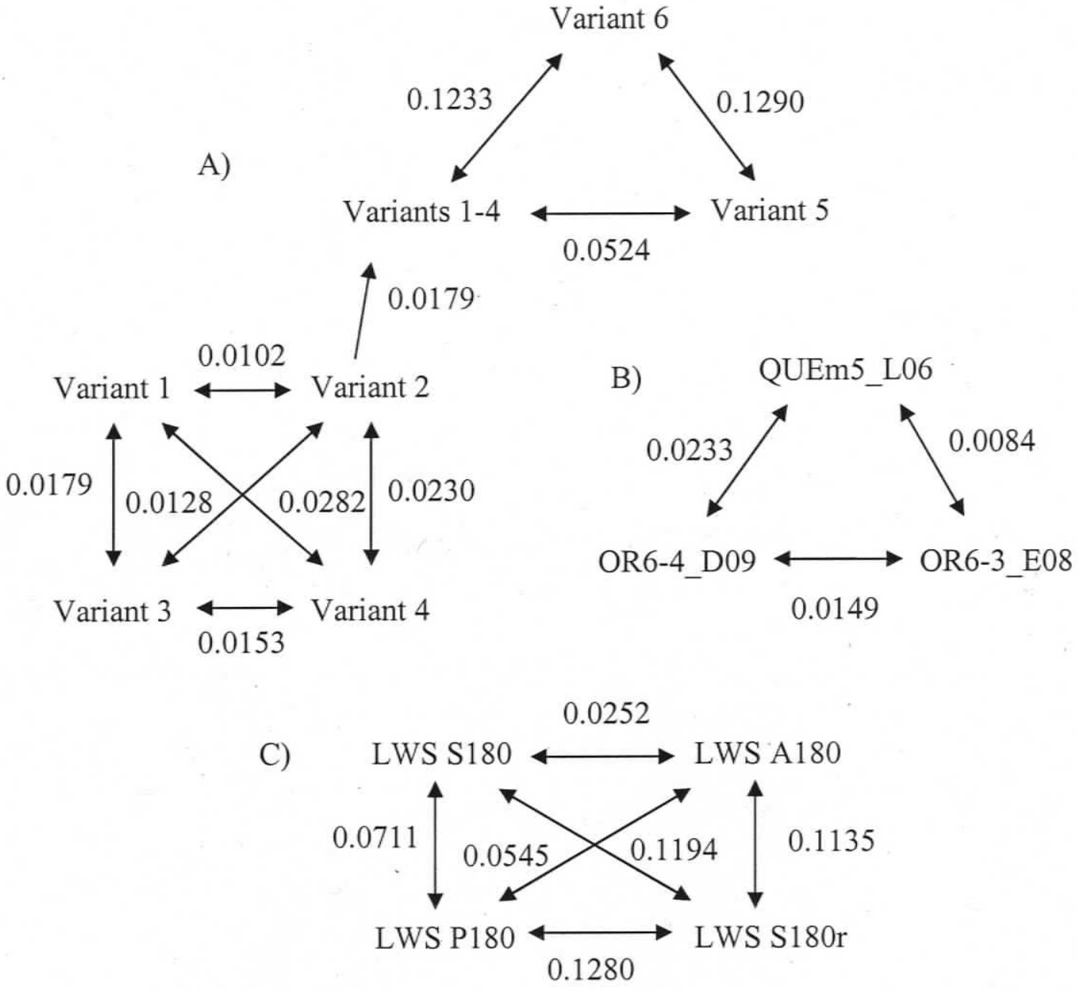


Figure 2.8. *Pairwise p-distance of guppy LWS opsins presented by Weadick and Chang (2007), Hoffman et al. (2007), and Ward et al. (2007).* (A) Weadick and Chang (2007) variants 1 – 6. Variants 1 – 2 and variants 3 – 4, posses the smallest p-distance within the variant 1 – 4 group, and may thus represent alleles at either the LWS S180 or LWS A180 loci. Variant 5 and variant 6 have a large p-distance from each other and to the grouped mean p-distance of variants 1 – 4. This suggests that both variants 5 and 6 likely represent separate loci. (B) Hoffman et al. (2007) OR6-4_D09, OR6-3_E08, QUEm5_L06 sequences. The p-distance between QUEm5_L06 and OR6-3_E08 is very small suggesting that these sequences represent alleles at the same genetic locus. The p-distance of OR6-4_D09 to OR6-3_E08 and QUEm5_L06 is large enough that this sequence may represent a separate genomic locus. (C) Ward et al. (2008) LWS S180, LWS A180, LWS P180 and LWS S180r sequences. These sequences represent the largest p-distances of the three studies and in turn, likely represent the four LWS opsin loci present in the Cumaná guppy.

Section 4: Expressional Analysis and Future Directions

At the genetic level, the guppy and species of the family Poeciliidae, possess a greater complement of LWS opsin genes than any other vertebrate investigated to this time. The guppy also appears to have population-level variation driven by gene conversion; perhaps facilitating differential spectral acuities among its populations. Although a proper genomic characterization of these LWS paralogs both between and among guppy populations is important (see above section); the notion that these duplicates could provide an increased spectral sensitivity to red/orange light relies on the assumption that all paralogs are expressed simultaneously. Therefore, in the fourth and final section of chapter 2, LWS opsin paralog expression within the Cumaná guppy retina will be discussed. Furthermore, important future directions using retinal *in situ* hybridization and behavioural testing will be outlined to help complete the story of colour detection ability within this species and others in Poeciliidae.

LWS opsin paralog expression in the guppy retina

As discussed in chapter 1, RT-PCR was used to confirm the presence of all four LWS transcripts in a single guppy adult eye. That is, transcripts for LWS S180, LWS A180, LWS P180 and LWS S180r were each amplified from a cDNA template and sequenced to confirm identity. To verify that the intronless LWS S180r paralog was in fact expressed and not an artefact of genomic contamination during RT-PCR; it was observed that the cDNA template used to amplify this paralog did not co-amplify genomic DNA while detecting the other LWS transcripts. This observation suggests that the DNase added during the RNA isolation completely removed all genomic DNA. Further to this point, Weadick and Chang (2007) also amplified the LWS S180r transcript (called 'variant 6' in their own study) from a cDNA template; adding further evidence that this intronless paralog is indeed expressed. Knowing that all four LWS transcripts were expressed in the retina of the adult Cumaná guppy (and likely the Paria River guppy as demonstrated by Weadick and Chang, 2007), we were then interested in determining how early in development any of these LWS paralogs could be detected. Of interest, we were able to amplify both the LWS S180 and LWS P180 transcripts from a single, late-eyed (Martyn et al., 2006) embryo. To our knowledge, this represents the first time that expressed LWS opsin paralogs have been identified during embryonic development in a live-bearing fish. These results may lead to future investigations with regards to developmental spatial and temporal patterns in guppy LWS expression; much like the ISH work already performed by Takechi and Kawamura (2005) on the developing zebrafish retina (see below).

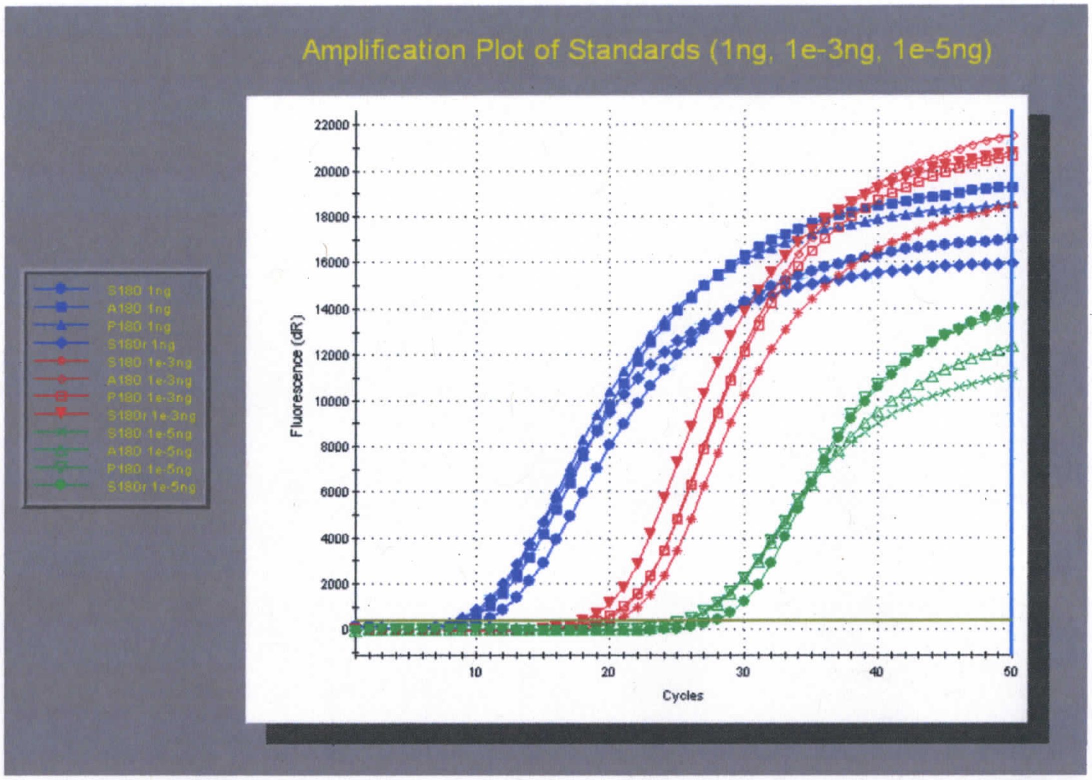
Complete reading frames, expected GPCR configuration (Weadick and Chang, 2007), RT-PCR data presented here, and MSP results from Archer et al. (1987, 1990); suggests that the LWS transcripts identified in this study are indeed translated into functional opsins. Of importance, however, the spectral influence that each LWS duplicate has on colour detection ability in Poeciliidae should be examined more carefully. As shown in table 1.3, λ_{max} predictions for the LWS S180, LWS A180 and

LWS S180r transcripts are possible using the five key-sites rule proposed by Yokoyama and Radlwimmer (2001). The predicted spectral sensitivity of these three LWS transcripts, however, does not correlate precisely with the long-wavelength spectral range in guppies observed by Archer et al. (1987, 1990); parcticularly those in the shorter wavelength visual pigments (~529 nm). It is perhaps the LWS P180 transcript, with key-site substitutions suggesting a shorter-wavelength sensitive opsin, which may be contributing to the observed λ_{max} shift in some guppy visual pigments. By constructing this LWS opsin and evaluating its absorption spectra through an *in vitro* assay (Yokoyama and Radlwimmer, 2001), it would be possible to determine not only the spectral influence of a proline residue at amino acid position 180 (see chapter 1), but also the spectral absorption curve of this LWS paralog. A similar experiment could be done on each LWS transcript of a given population; whereby the total area under the absorptions curves would determine an accurate spectral potential for the tested individuals. It is important to note, however, that the spectral range of opsins in general is not governed only by key-site amino acids, but also by the ratios of rhodopsin (A1) and porphyropsin (A2) used as vitamin A derived chromophores. Therefore, determining the exact spectral sensitivity of each LWS paralog in relation to colour detection ability for this species, must also rely on the A1/A2 ratios present in the retinal cells of the tested fish.

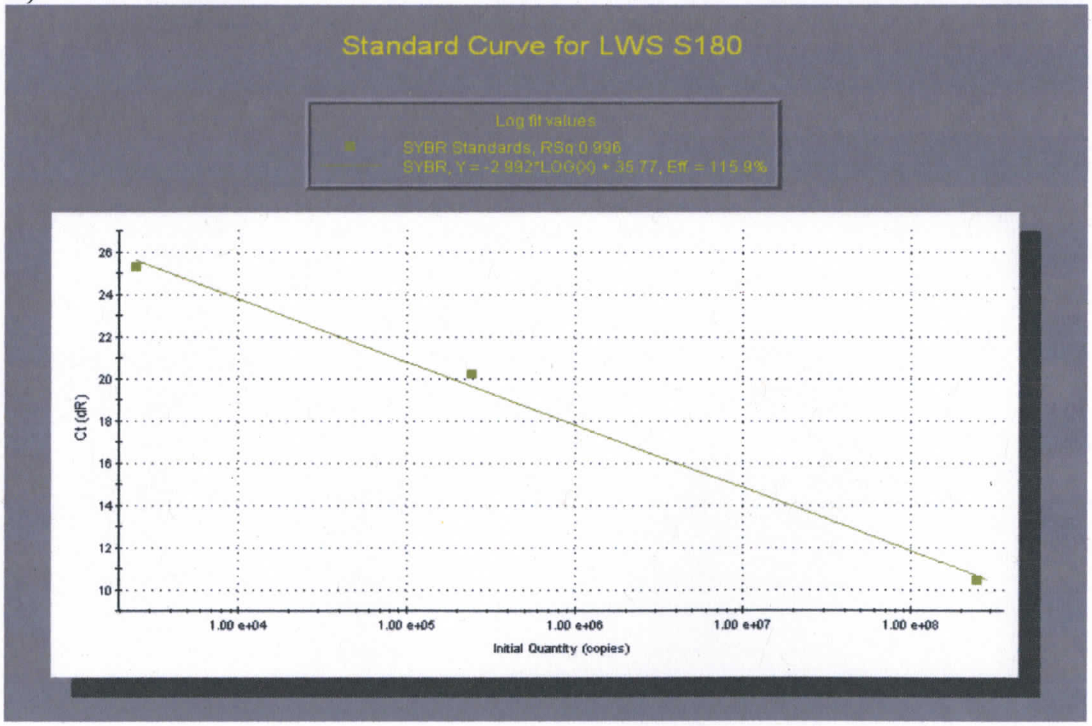
Quantification LWS opsin paralog expression in the guppy retina

We determined the amount of duplicate expression in three Cumaná guppy individuals using quantitative PCR (qPCR) on cDNA derived from adult retinas. qPCR is a useful tool for determining the biological relevance of gene expression, as RT-PCR can only identify a gene as being expressed and says nothing about its relative abundance. The detailed methodology of this experiment is described in the methods section of chapter 1, although the standard amplification plot, standard curves and dissociation plot is presented here (see figure 2.9). From these plots and the data presented in figure 1.4, we were able to calculate the initial copy-number of LWS paralogs in our diluted cDNA samples. From this, we determined that LWS A180 was highly expressed in these three individuals and that LWS S180 and LWS S180r were also expressed, albeit at lower levels. Finally, LWS P180 appeared to be expressed very minimally (see figure 1.4) which signified a surprising result given that this transcript is suspected to contribute to the shorter-wave visual pigments of the guppy's retina, as observed by Archer et al. (1987, 1990).

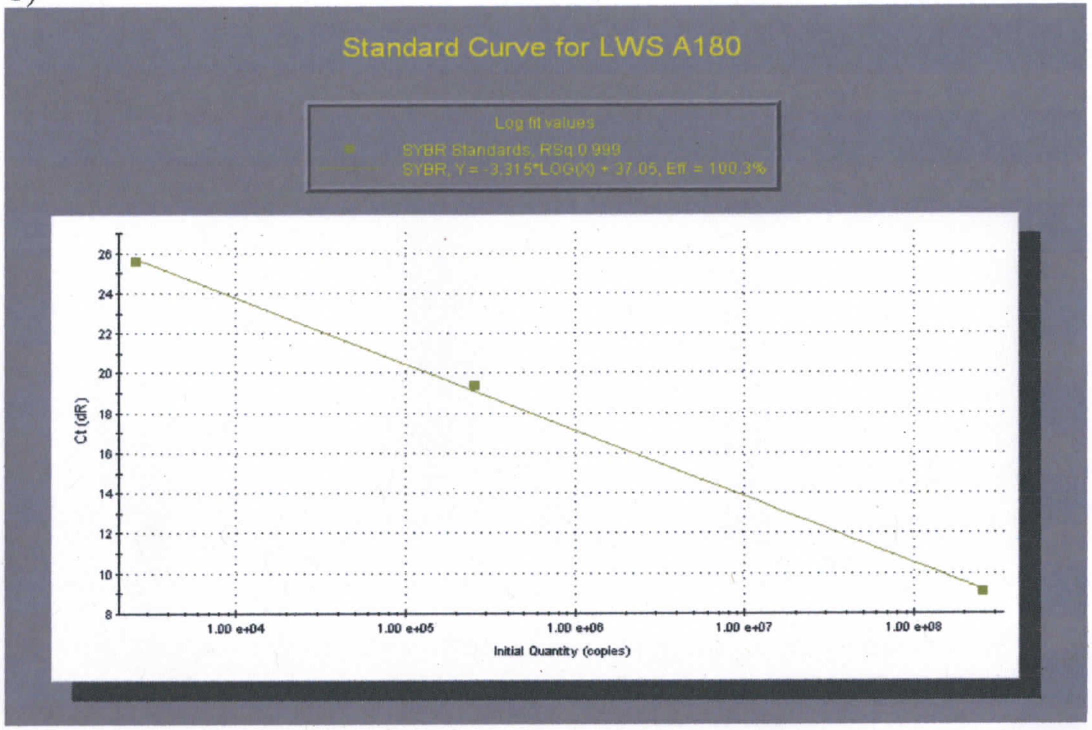
A)


Amplification Plot of Standards (1ng, 1e-3ng, 1e-5ng)
S180 1ng
A180 1ng
P180 1ng
S180r 1ng
S180 1e-3ng
A180 1e-3ng
P180 1e-3ng
S180r 1e-3ng
S180 1e-5ng
A180 1e-5ng
P180 1e-5ng
S180r 1e-5ng
Fluorescence (dR)
22000
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
10
20
30
40
50
Cycles

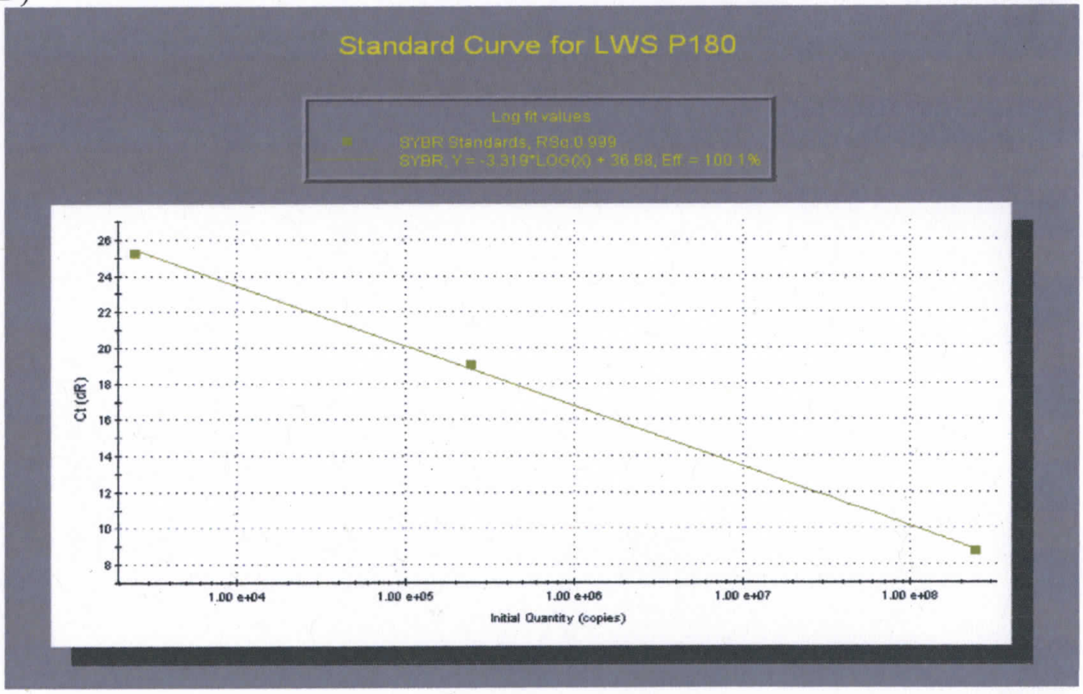
B)


Standard Curve for LWS S180
Log fit values
SYBR Standards, RSq:0.996
SYBR, Y = -2.992*LOG(X) + 35.77, Eff. = 115.9%
Ct (dR)
26
24
22
20
18
16
14
12
10
1.00 e+04
1.00 e+05
1.00 e+06
1.00 e+07
1.00 e+08
Initial Quantity (copies)

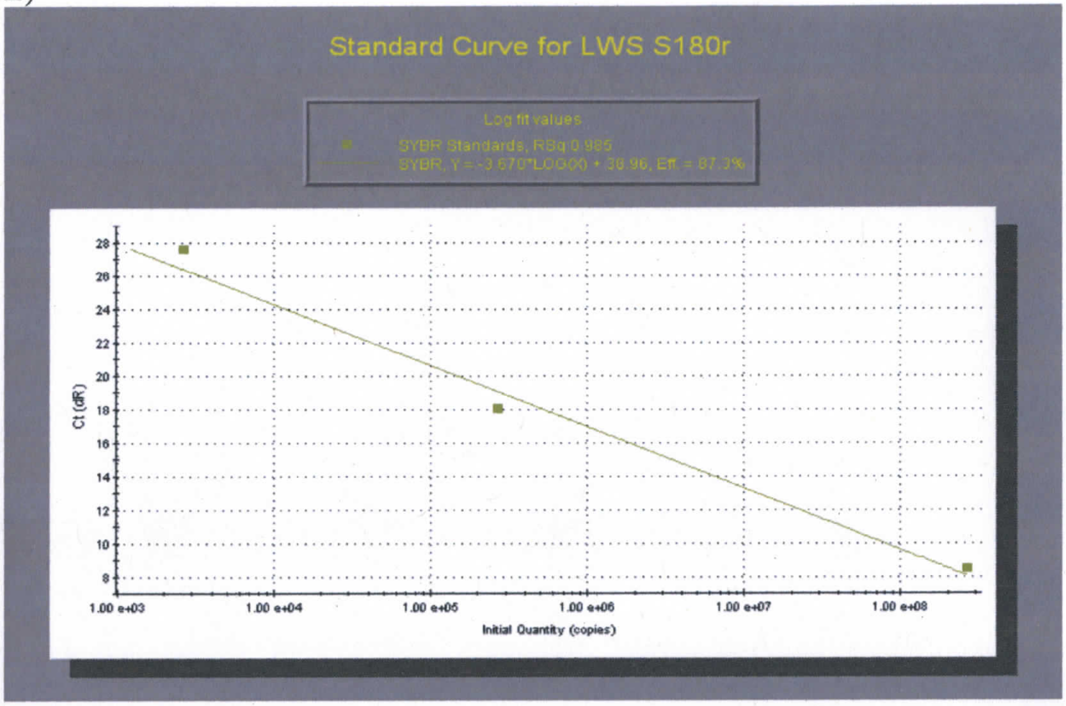
C)


Standard Curve for LWS A180
Log fit values
SYBR Standards, RSq:0.999
SYBR, Y = -3.315*LOG(X) + 37.05, Eff. = 100.3%
Ct (dR)
26
24
22
20
18
16
14
12
10
8
1.00 e+04
1.00 e+05
1.00 e+06
1.00 e+07
1.00 e+08
Initial Quantity (copies)

D)


Standard Curve for LWS P180
Log fit values
SYBR Standards, RSq:0.999
SYBR, Y = -3.319*LOG(X) + 36.68, Eff. = 100.1%
Ct (dR)
26
24
22
20
18
16
14
12
10
8
1.00 e+04
1.00 e+05
1.00 e+06
1.00 e+07
1.00 e+08
Initial Quantity (copies)

E)


Standard Curve for LWS S180r
Log fit values
SYBR Standards, RSq:0.985
SYBR, Y = -3.670*LOG(X) + 38.96, Eff. = 87.3%
Ct (dR)
28
26
24
22
20
18
16
14
12
10
8
1.00 e+03
1.00 e+04
1.00 e+05
1.00 e+06
1.00 e+07
1.00 e+08
Initial Quantity (copies)

F)

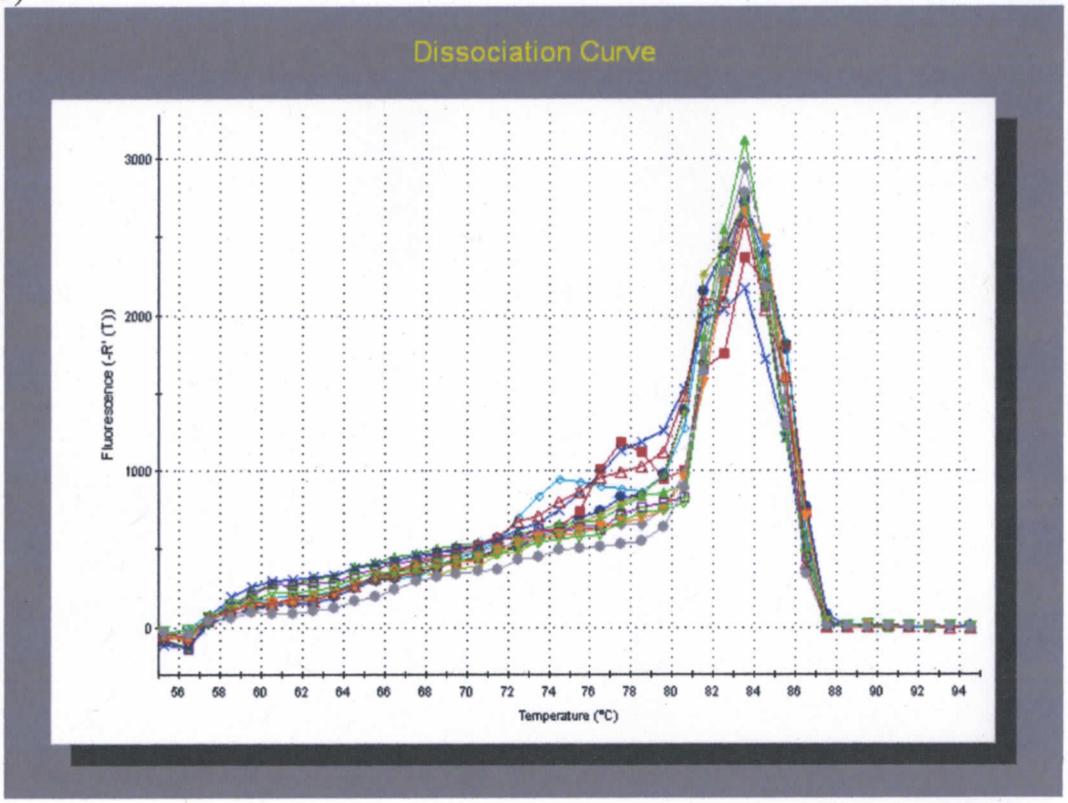

Dissociation Curve
Fluorescence (-R'(T))
3000
2000
1000
0
56
58
60
62
64
66
68
70
72
74
76
78
80
82
84
86
88
90
92
94
Temperature (°C)

Figure 2.9. *Amplification plot of standards; standard curves; and qPCR dissociation plot.* A) Amplification plot of standards. For LWS S180, LWS A180, LWS P180 and LWS S180r, 1ng of plasmid-inserted template contained approximately 2.447×10^8, 2.569×10^8, 2.454×10^8, and 2.669×10^8 initial copies, respectively. B-E) Standard curves of Ct(dR) values compared to initial copy number is shown for LWS S180, LWS A180, LWS P180 and LWS S180r (see table 1.5). F) Dissociation curve of Fluorescence [–R'(T)] over T°C during quantitative PCR of all four cDNA derived transcripts. Individual curves are not labeled, yet each show dissociation around 82-85°C; suggesting that co-amplification of contaminants (i.e. primer dimers) did not occur during qPCR. This prediction was further confirmed using gel electrophoresis of qPCR amplicons (data not shown).

Using quantitative PCR to determine the relative amount of LWS paralog expression in the Cumaná guppy represents an important investigation, as the methodology may now be used to repeat this experiment on a much larger sample size of guppies and Poeciliidae fish. As mentioned above, Archer et al. (1987, 1990) have shown that the spectral sensitivity of long-wavelength cones in this species is highly variable, with both males and females exhibiting populations of photopigments in the 529-579nm range. However, the demonstrated trend of LWS paralog expression shown here through qPCR does not support the spectral range of visual pigments observed in these fish. This is parcticularly evident as the LWS A180, LWS S180 and LWS S180r transcripts could almost certainly not embody any of the shorter-wavelength visual pigments observed in many guppy individuals. It is thus likely that a larger sample size of multiple individuals and populations would reveal polymorphisms in LWS expression (parcticularly in regards to transcription of LWS P180) – either among individuals or possibly during development as is the case in the zebrafish (Takechi and Kawamura, 2005). It is therefore an interesting notion that both gene conversion (see above) among populations, and the potential for variable expression of LWS paralogs, may have facilitated the disparity in colour detection observed among and within populations of this species.

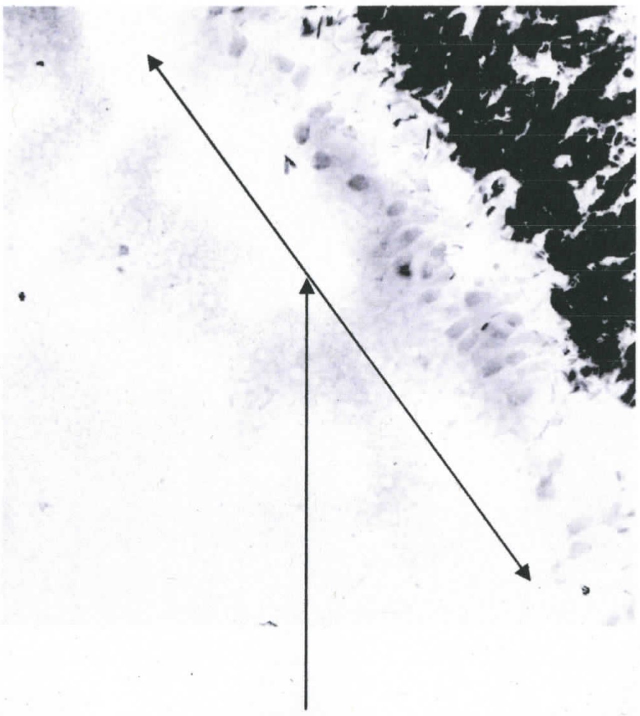
Ecological implications of variable spectral tuning for the long-wavelength visual region

Variable spectral tuning, governed by the above factors may be important to these fish for numerous reasons. For instance, water turbidity, which differs among guppy populations, may impact a female's ability to select a mate for copulation, and in turn, necessitate the need for spectral tuning to visualize the variable contrasts in male morphological patterning. Indeed, many animals increase the probability that their own visual signals are detected by evolving conspicuous traits to contrast with their visual backgrounds (Uy and Endler, 2004). Thus, the various surroundings that multiple populations of the guppy exist in may necessitate the need for individual spectral variation and in turn, allow for proper mate selection via female choice criteria. Other factors, however, to evolve differential spectral acuities in the guppy may include predator avoidance and foraging efficiency, which also varies within the multiple populations of this species. Here, gene conversion and expression variation of LWS paralogs may allow for tuning of visual acuity to best balance and equalize these important biological factors. Indeed, White et al. (2005) found that blocking of long wavelength light significantly reduced guppy foraging rates and in turn, contributed more to prey detection than any other wavelength. By varying spectral acuities both between and within populations, the guppy may be able to adapt to changing conditions in foraging potential and predator avoidance, thus increasing the overall fitness of the species. Finally, plasticity in spectral tuning may also have helped to facilitate speciation within Poeciliidae as differential colour detection has perhaps allowed for the divergence of these species. Regardless of the rationale, it is clear that along with an increased capacity for red/orange spectral acuity, variation in colour detection exists within the guppy species (and possibly within Poeciliidae), and is likely facilitated at both the genetic and expressional levels. Future investigations in LWS transcript variation should therefore include a much larger sample size of tested individuals, and perhaps, a more inclusive analysis of other Poeciliidae fish to determine if a causative link to speciation may exist.

Preliminary in situ hybridization results

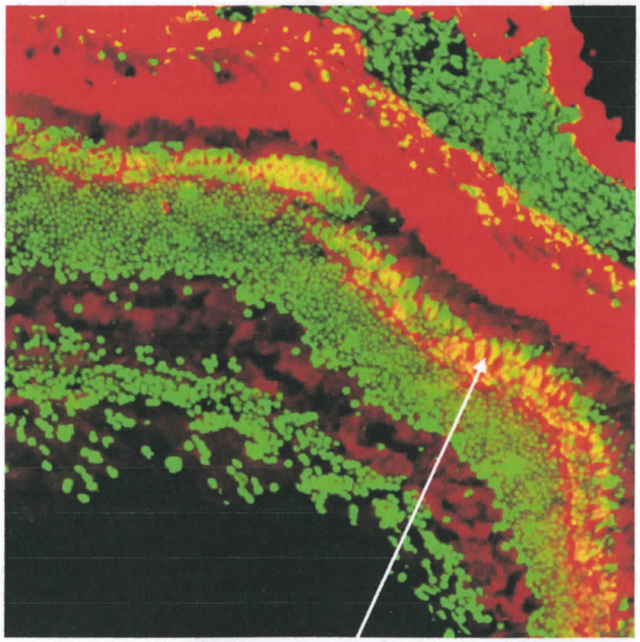
As mentioned above and in chapter 1, the use of paralogous opsins to increase spectral acuity relies not only on transcript abundance, but on temporal and spatial patterns of expression within the retina of a species. For instance, Tackechi and Kawamura (2005) have used *in situ* hybridization to show a dynamic temporal and spatial pattern of opsin subtype expression in the zebrafish retina. Because the Cumaná guppy has been characterized to have a larger repertoire of expressed LWS duplicates than any other vertebrate, we felt this species would be a prime candidate to investigate paralogous LWS opsin expression within the adult retina. Unfortunately, LWS paralog identity is too high within the coding region of these paralogs for proper differentiation between target probes. We therefore attempted to examine paralog expression in the Cumaná guppy adult retina using two DIG-labeled oligonucleotide probes targeted for sequence available 3'UTR. The probes were designed and synthesized by GeneDetect$^{\mathrm{TM}}$, limited, with one probe targeting the 3'UTR of the LWS S180 transcript, and the other targeting the 3'UTR of the LWS A180 and LWS P180 transcripts (which are identical in this region due to the hybrid nature of the LWS A180 locus). Unfortunately, this experiment was met with inconclusive results, as only the LWS S180 probe showed any signal of proper hybridization. These results are shown below in figure 2.10, with transcript signaling present throughout the entire dorsal to ventral photoreceptor layer. Some background signaling was also observed in the surrounding tissue, suggesting that the specificity of this probe may be in question; parcticularly as it consisted of a high (62.5%) A/T nucleotide content. However, this experiment has set the way for future *in situ* hybridization work on the guppy retina, as plans are currently underway to design full-length RNA probes for each of the LWS duplicates in the hopes of characterizing the spatial and temporal patterns of retinal paralog expression.

A)



LWS S180 signal

B)



Photoreceptor layer:
LWS S180 and
bis-benzamide signals

Figure 2.10. *In situ hybridization with an LWS S180 probe in the retina of two adult female Cumaná guppies.* A) *In situ* hybridization of an LWS S180 probe. Shown is a 16μm retinal cross section (dorsal to ventral plane) cut through the optic nerve (not shown in figure) and hybridized with a 48mer GreenStar^{TM} DIG-labeled oligonucleotide probe; targeted to the 3'UTR of the LWS S180 transcript and synthesized from GeneDetect^{TM} Limited. Protocol used was modified from the GeneDetect^{TM} laboratory methods for ISH using oligonucleotide probes. Hybridization of the probe was performed overnight at room temperature, washed using 1XSSC at 55°C and detected using anti-DIG antibodies and NBT/BCIP tablets from Roche Applied Science^{TM}. B) Composite image of ISH for LWS S180 and bis-benzamide (Fisher^{TM}) labeled nuclei. ISH using the above protocol was repeated on second female fish and stained with bis-benzamide to identify retinal nuclei. The image is a red/green inversion, with yellow composite within the photoreceptor layer showing overlap between the LWS S180 signal and its corresponding stained nuclei.

Future directions using behavioural testing

Behavioural experiments have been developed to look for correlations between opsin gene sequence variation and visual performance. In humans, it appears that an extra opsin gene (in females that are heterozygous for either LWS or MWS) improves wavelength discrimination (Jameson and Highnote, 2001). Muntz and Cronly-Dillon (1966) correlated photoreceptor variation (based upon MSP data) with functional variation by training goldfish to discriminate blue from green, blue from red, and red from green in food trough experiments. Another technique used to measure wavelength sensitivity in fish is the optomotor response in which fish follow a pattern of moving stripes. No training is required and this behavior is observed in species from a diversity of evolutionary lineages. By measuring the ability of guppies to follow different wavelengths of light, Anstis et al. (1998) determined that guppies have photoreceptors sensitive to red, green, and blue light. Similar experiments could be used to more precisely measure wavelength discrimination in guppies.

The hypothesis that sexual selection in guppies was facilitated by the evolution of multiple LWS loci does not mean that these events occur at the same time. Indeed colour-based mate choice appears to have evolved in only a subset of poeciliids with extra LWS opsin genes. In these species, male colouration might be exploiting a trait that is an adaptation to prey detection. As discussed above, White et al. (2005) found that blocking long wavelength light significantly reduced guppy foraging rates. Thus, female guppies, through LWS opsin gene duplication and divergence, may be pre-adapted to both maximize foraging efficiency and to evaluate male colour variation. Thus, colour detection ability in this species likely plays an important role in the multiple behaviors exhibited by these fish and behavioural tests to demonstrate the importance of an enhanced acuity in the long-wavelength spectral region should be of the utmost importance in future investigations.

Thesis Conclusion

Gene duplication and divergence has provided *Poecilia* and its close relatives with four distinct LWS opsin genes; a larger repertoire than any other vertebrate. Furthermore, our phylogenetic analyses suggest that three of these LWS opsins (LWS S180r, LWS P180 and LWS S180) occur in at least 239 species of the subfamily Poeciliinae. Adult guppies express all four LWS paralogs simultaneously, albeit at varying levels. As a consequence of these gene duplications, enhanced wavelength discrimination in the long-wave portion of the visible spectrum has likely facilitated a red-orange colour bias for sexual selection within the guppy and possibly its sister taxa, Micropoecilia.

References

Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., Polz, M.F. 2005. PCR-induced sequence artefacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Applied and Environmental Microbiology. 71(12), 8966-8969.

Alexander, H.J., Breden, F. 2004. Sexual isolation and extreme morphological divergence in the Cumaná guppy: a possible case of incipient speciation. Journal of Evolutionary Biology. 17, 1238-1254.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology. 215(3), 403-410.

Andersson, M. 1994. Sexual Selection. Princeton University Press. Princeton, New Jersey. pp. 599.

Anstis, S.M., Hutahajan, P., Cavanagh, P. 1998. Optomotor test for wavelength sensitivity in guppufish (Poecilia reticulata). Vision Research. 38, 45-53.

Archer, S.N., Endler, J.A., Lythgoe, J.N., Partridge, J.C. 1987. Visual pigment polymorphism in the guppy, *Poecilia reticulata*. Vision Research. 27, 1243-1252.

Archer, S.N., Lythgoe, J.N. 1990. The visual pigment basis for cone polymorphism in the guppy, *Poecilia reticulata*. Vision Research. 30, 225-233.

Archer, S., Hope, A., Partridge, J.C. 1995. The molecular basis for the green-blue sensitivity shift in the rod visual pigments of the European eel. Proceedings of the Royal Society London B. 262, 289-295.

Ashelford, K.E., Chuzhanove, N.A., Fry, J.C., Jones, A.J., Weightman, A.J. 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Applied and Environmental Microbiology. 71(12), 7724-7736.

Bakker, T.C.M. 1993. Positive genetic correlation between female preference and preferred male ornament in sticklebacks. Nature. 363, 255-257.

Bergsten J. 2005. A review of long-branch attraction. Cladistics. 21, 163-193.

Bourne, G.R., Breden, F., Allen, T.C. 2003. Females prefer carotenoid males as mates in the pentamorphic livebearing fish, *Poecilia parae*. Naturwissenschaften. 90, 402-405.

Bradly, R.D., Hillis, D.M. 1997. Recombinant DNA sequences generated by PCR amplification. Molecular Biology and Evolution. 14(5), 592-593.

Breden, F., Stoner, G. 1987. Male predation risk determines female preference in the Trinidad guppy. Nature. 329, 831-833.

Breden, F., Ptacek, M.B., Rashed, M., Taphorn, D., Figueiredo, C.A. 1999. Molecular phylogeny of a live-bearing fish genus *Poecilia* (Poeciliidae: Cyprinodontiformes). Molecular Phylogenetics and Evolution. 12, 95-104.

Briscoe, A.D. 2001. Functional diversification of lepidopteran opsins following gene duplication. Molecular Biology and Evolution. 18(12), 2270-2279.

Brooks, R. 2002. Variation in female mate choice within guppy populations: population divergence, multiple ornaments and the maintenance of polymorphism. Genetica. 116, 343-358.

Carelton, K.L., Kocher, T.D. 2001. Cone opsin genes of African cichlid fishes: Tuning spectral sensitivity by differential gene expression. Molecular Biology and Evolution. 18(8), 1540-1550.

Carleton, K.L., Parry, J.W.L., Bowmaker, J.K., Hunt, D.M., Seehausen, O. 2005. Colour vision and speciation in Lake Victoria cichlids of the genus *Pundamilia*. Molecular Ecology. 14, 4341-4353.

Carroll, J. 2007. Focus on molecules: the cone opsins. Experimental Eye Research. In Press.

Cepko, C.L. 1996. The patterning and onset of opsin expression in vertebrate retinae. Current Opinion in Neurobiology. 6, 542-546.

Chen, J-M., Chuzhanova, N., Stenson, P.D., Férec, C., Cooper, D.N. 2005. Intrachromosomal serial replication slippage in trans gives rise to diverse genomic rearrangements involving inversion. Human Mutation. 26, 362-373.

Chinen, A., Hamaoka, T., Yamade, Y., Kawamura, S. 2003. Gene duplication and spectral diversification of cone visual pigments of Zebrafish. Genetics. 163, 663-675.

Chinen, A., Matsumoto, Y., Kawamura, S. 2005. Reconstruction of ancestral green visual pigments of zebrafish and molecular mechanism of their spectral differentiation. Molecular Biology and Evolution. 22, 1001-1010.

Churcher, A.M., Taylor, J.S. 2008. A review of PCR medicated template switching. In revision.

Crispo, E., Bentzen, P., Reznick, D.N., Kinnison, M.T., Hendry, A.P. 2006. The relative influence of natural selection and geography on gene flow in guppies. Molecular Ecology. 15, 49-62.

Cronn, R., Cedroni, M. Haselkorn, T., Grover, C., Wendel, J.F. 2002. PCR-mediated recombination in amplification products derived from polyploidy cotton. Theoretical Applied Genetics. 104, 482-489.

Davies, W.L., Cowing, J.A., Carvalho, L.S., Potter, I.C., Trezise, A.E.O., Hunt, D.M., Collin, S.P. 2007. Functional characterization, tuning, and regulation of visual pigment gene expression in an anadromous lamprey. FASEB Journal. 21, 2713-2724.

Deeb, S.S. 2004. Molecular genetics of colour-vision deficiencies. Visual Neuroscience. 21, 191-196.

Dulai, K.S., von Dornum, M., Mollon, J.D., Hunt, D.M. 1999. The evolution of trichromatic colour vision by opsin gene duplication in New World and Old World primates. Genome Research. 9, 401-410.

Endler, J.A. 1980. Natural selection on colour patterns in *Poecilia reticulata*. Evolution. 34, 76-91.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 39, 783-791.

Godin, J.G., Dugatkin, L.A. 1996. Female mating preference for bold males in the guppy, *Poecilia reticulata*. Proceedings of the National Academy of Science USA. 93(19), 10262-10267.

Grether, G.F., Hudon, J., Endler, J.A. 2001. Carotenoid scarcity, synthetic pteridine pigments and the evolution of sexual coloation in guppies (*Poecilia reticulata*). Proceedings of the Royal Society of London B. 268, 1245-1253.

Grether, G.F., Kolluru, G.R., Rodd, H.F., Cerda, J., Shimazaki, K. 2005. Carotenoid availability affects the development of a colour-based mate preference and the sensory bias to which it is genetically linked. Proceedings of the Royal Society. 272, 2181-2188.

Gunther, K.L., Neitz, J., Neitz, M. 2006. A novel mutation in the short-wavelength-sensitive cone pigment gene associated with a tritan colour vision defect. Visual Neuroscience. 23, 403-409.

Haesler, M.P., Seehausen, O. 2005. Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: implications for speciation. Proceedings of the Royal Society B. 272, 237-245.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41, 95-98.

Halstenberg, S., Lindgren, K.M., Samagh, S.P.S., Nadal-Vicens, M., Balt, S., Fernald, R.D. 2005. Diurnal rhythm of cone opsin expression in the teleost fish *Haplochromis burtoni*. Visual Neuroscience. 22, 135-141.

Handley, L.J.L., Manica, A., Goudet, J., Balloux, F. 2007. Going the distance: human population genetics in a clinal world. Trends in Genetics. 23(9), 432-439.

Hoffmann, M., Tripathi, N., Henz, S.R., Lindholm, A.K., Weigel, D., Breden, F., Dreyer, C. 2007. Opsin gene duplication and diversification in the guppy, a model for sexual selection. Proceedings of the Royal Society London B. 274, 33-42.

Hoke, K.L., Evans, B.I., Fernald, R.D. 2006. Remodeling of the cone photoreceptor mosaic during metamorphosis of flounder (*Psuedopleuronectes americanus*). Brain Behavior and Evolution. 68, 241-254.

Houde, A.E. 1988. Genetic difference in female choice between two guppy populations. Animal Behavior. 36(2), 510-516.

Houde, A.E. 1997. Sex, Colour, and Mate Choice in Guppies. Princeton University Press. Princeton, New Jersey. pp. 210.

Houde, A.E. and Endler, J.A. 1990. Correlated evolution of female mating preferences and male colour patterns in the guppy *Poecilia reticulata*. Science. 248, 1405-1408.

Houde, A.E., Hankes, M.A. 1997. Evolutionary mismatch of mating preferences and male colour patterns in guppies. Animal Behavior. 53, 343-351.

Hrbek, T., Seckinger, J., Meyer, A. 2007. A phylogenetic and biogeographic perspective on the evolution of poeciliid fishes. Molecular Phylogenetics and Evolution. 43, 986-998.

Hugenholtz, P., Huber, T. 2003. Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. International Journal of Systematic and Evolutionary Microbiology. 53, 289-293.

Hughes, K.A., Du, L., Rodd, F.H., Reznick, D.N. 1999. Familiarity leads to female mate preference for novel males in the guppy, *Poecilia reticulata*. Animal Behavior. 58, 907-916.

Hyrien, O., Debatisse, M., Buttin, G., de Saint Vincent, B.R. 1988. The multicopy appearance of a large inverted duplication and the sequence at the inversion joint suggests a new model for gene amplification. The EMBO Journal. 7, 407-417.

Ishii, K., Fukui, M. 2001. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. Applied and Enviromental Microbiology. 67(8), 3753-3755.

Iyer, R.R., Pluciennik, A., Burdett, V., Modrich, P.L. 2006. DNA mismatch repair: functions and mechanisms. Chemical Reviews. 106, 302-323.

Jacobs, G.H., Williams, G.A., Cahill, H., Nathans, J. 2007. Emergence of novel colour vision in mice engineered to express a human cone photopigment. Science. 315, 1723-1725.

Jagla, W.M., Jägle, H., Hayashi, T., Sharpe, L.T., Deeb, S.S. 2002. The molecular basis of dichromatic vision in males with multiple red and green visual pigment genes. Human Molecular Genetics. 11(1), 23-32.

Jameson, K.A., Highnote, S.M. 2001. Richer colour experience in observers with multiple photopigment opsin genes. Psychonomic Bulletin and Review. 8, 244-261.

Jones, M.P., Pierce, K.E., Ward, D. 2007. Avian vision: a review of form and function with special consideration to birds of prey. Journal of Exotic Pet Medicine. 16(2), 69-87.

Judo, M.S.B., Wedel, A.B., Wilson, C. 1998. Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Research. 26(7), 1819-1825.

Jukes, T.H., Cantor. C.R. 1969. Evolution of protein molecules. Pp. 21-123 in H. N. Munro, ed. *Mammalian protein metabolism*. Academic Press, New York.

Körner, K.E., Schlupp, I., Plath, M., Loew, E.R. 2006. Spectral sensitivity of mollies: comparing surface- and cave-dwelling Atlantic mollies, *Poecilia mexicana*. Journal of Fish Biology. 69, 54-65.

Kumar, S., Tamura, K., Nei, M. 2004. MEGA4: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics. 5, 150-163.

Kunkel, T.A., Erie, D.A. 2005. DNA mismatch repair. Annual Review of Biochemistry. 74, 681-710.

Kunz, Y.W., Ennis, S., Wise, C. 1983. Ontogeny of the photoreceptors in the embryonic retina of the viviparous guppy, *Poecilia reticulata P.* (Teleostei). Cell Tissue Research. 230, 469-486.

Kurata, S., Kanagawa, T., Magariyama, Y., Takatsu, K., Yamada, K., Yokomaku, T., Kamagata, Y. 2004. Reevaluation and reduction of a PCR bias caused by reannealing of templates. Applied and Environmental Microbiology. 70(12), 7545-7549.

L'Abbé, D., Belmaaza, A., Décary, F., Chartrand, P. 1992. Elimination of heteroduplex artefacts when sequencing HLA genes amplified by polymerase chain reaction (PCR). Immunogenetics. 35, 395-397.

Lennie, P. 2000. Colour vision: putting it together. Current Biology. 10, R589-R591.

Li, K-B. 2003. ClustalW-MPI: ClustalW analysis using distributed and parallel computing. Bioinformatics. 19, 1585-1586.

Li, P., Temple, S., Gao, Y., Haimberger, T.J., Hawryshyn, C.W., and Li, L. 2005. Circadian rhythms of behavioural cone sensitivity and long wavelength opsin mRNA expression: a correlation study in zebrafish. Journal of Experimental Biology. 208, 497-504.

Lindholm, A., Brooks, R., Breden F. 2004. Extreme polymorphism in a Y-linked, sexually selected trait. Heredity. 92, 156-162.

Longeri, M., Zanotti, M., Damiani, G. 2002. Recombinant DRB sequences produced by mismatch repair of heteroduplexes during cloning in *Escherichia coli*. European Journal of Immunogenetics. 29, 517-523.

Lotto, R.B., Purves, D. 2002. The empirical basis of colour perception. Consciousness and Cognition. 11, 609-629.

Luyten, P.H., Liley, N.R. 1985. Geographic variation in the sexual behavior of the guppy, *Poecilia reticulata* (Peters). Behavior. 95, 164-179.

Maan, M.E., Hofker, K.D., van Alphen, J.J.M., Seehausen, O. 2006. Sensory drive in cichlid speciation. The American Naturalist. 167, 947-954.

Macke, J.P., Nathans, J. 1997. Individual variation in the size of the human red and green visual pigment gene array. Investigative Ophthalmology and Visual Science. 38(5), 1040-1043.

Masland, R.H. 2001. The fundamental plan of the retina. Nature Neuroscience. 4, 877-886.

Martyn, U., Weigel, D., Dreyer, C. 2005. *In vitro* culture of embryos of the guppy, *Poecilia reticulata*. Developmental Dynamics. 235, 617-622.

Matsumoto, Y., Fukamachi, S., Mitani, H., Kawamura, S. 2006. Functional characterization of visual opsin repertoire in Medaka (*Oryzias latipes*). Gene. 371, 268-278.

Meyer, A., Lydeard, C. 1994. The evolution of copulatory organs, internal fertilization, placentae and viviparity in killifishes (Cyprinodontiformes) inferred from a DNA phylogeny of the tyrosine kinase gene *X-src*. Proceedings of the Royal Society London Series B. 254, 153-162.

Meyerhans, A., Vartanian, J.P., Wain-Hobson, S. 1990. DNA recombination during PCR. Nucleic Acids Research. 18(7), 1687-1691.

Muntz, W.R.A., Cronly-Dillon, J.R. 1966. Colour discrimination in goldfish. Animal Behavior. 14(2), 351-355.

Nathans, J., Thomas, D., Hogness, D.S. 1986. Molecular genetics of human colour vision: The genes encoding blue, green, and red pigments. Science. 232, 193-232.

Nei, M., Kumar, S. 2000. Molecular evolution and phylogenetics. Oxford University Press. New York, New York.

Nelson, D.L., Cox, M.M. 2005. Lehninger, Principles of Biochemistry, 4th ed. W.H. Freeman and Company. New York, New York. pp. 966-971.

Nicoletto, P.F. 1993. Female sexual response to condition-dependent ornaments in the guppy, *Poecilia reticulata*. Animal Behavior. 46, 441-450.

Odelberg, S.J., Weiss, R.B., Hata, A., White, R. 1995. Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. Nucleic Acids Research. 23(11), 2049-2057.

Pakula, A.A., Sauer, R.T. 1989. Genetic analysis of protein stability and function. Annual Review of Genetics. 23, 289-310.

Passananti, C., Davies, B., Ford, M., Fried, M. 1987. Structure of an inverted duplication formed as a first step in a gene amplification event: implications for a model of gene amplification. The EMBO Journal. 6,1697-1703.

Pilastro, A., Mandelli, M., Gasparini, C., Dadda, M., Bisazza, A. 2007. Copulation duration, insemination efficiency and male attractiveness in guppies. Animal Behavior. 74, 321-328.

Pointer, M.A., Cheng, C-H., C., Bowmaker, J.K., Parry, J.W.L., Soto, N., Jeffery, G., Cowing, J.A., Hunt, D.M. 2005. Adaptations to an extreme environment: retinal organization and spectral properties of photoreceptors in Antarctic notothenoid fish. Journal of Experimental Biology. 208, 2363-2376.

Polz, M.F., Cavanaugh, C.M. 1998. Bias in template-to-product ratios in multitemplate PCR. Applied and Environmental Microbiology. 64(10), 3724-3730.

Qiu, X., Wu, L., Huang, H., McDonel, P.E., Palumbo, A.V., Tiedje, J.M., Zhou, J. 2001. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Applied and Environmental Microbiology. 67(2), 880-887.

Raymond, P., Barthel, L.K., Curran, G.A. 1995. Developmental patterning of rod and cone photoreceptors in embryonic zebrafish. Journal of Comparative Neurology. 359, 537-550.

Robinson, J., Schmitt, E.A., Dowling, J.E. 1995. Temporal and spatial patterns of opsin gene expression in zebrafish (*Danio rerio*). Visual Neuroscience. 12, 895-906.

Rosen, D.E., Bailer, R.M. 1963. The poeciliid fishes (Cyprinodontiformes), their structure, zoogeography and systematics. Bulletin of the American Museum of Natural History. 126, 1-176.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. 239, 487-491.

Saitou, N., Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 4(4), 406-425.

Salzburger, W., Niederstätter, H., Brandstätter, A., Berger, B., Parson, W., Snoeks, J., Sturmbauer C. 2006. Colour-assortative mating among populations of *Tropheus moorii*, a cichlid fish from Lake Tanganyika, East Africa. Proceedings of Biological Science. 273(1584), 257-266.

Shammas, F.V., Heikkilä, R., Osland, A. 2001. Fluorescence-based method for measuring and determining the mechanisms of recombination in quantitative PCR. International Journal of Clinical Chemistry and Diagnostic Laboratory Medicine. 304, 19-28.

Shuldiner, A.R., Nirula, A., Roth, J. 1989. Hybrid DNA artefact from PCR of closely related target sequences. Nucleic Acids Research. 17(11), 4409.

Sison-Mangus, M.P., Bernard, G.D., Lampel, J., Briscoe, A.D. 2006. Beauty in the eye of the beholder: the two blue opsins of lycaenid butterflies and the opsin gene-driven evolution of sexually dimorphic eyes. The Journal of Experimental Biology. 209, 3079-3090.

Small, R.L. 2004. Use of nuclear genes for phylogeny reconstruction in plants. Australian Systematic Botany. 17, 145-170.

Spady, T.C., Seehausen, O., Loew, E.R., Jordan, R.C., Kocher, T.D., Carleton, K.L. 2005. Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. Molecular Biology and Evolution. 22, 1412-1422.

Streelman, J.T., Albertson, R.C., Kocher, T.D. 2003. Genome mapping of the orange blotch pattern in cichlid fishes. Molecular Ecology. 12, 2465-2471.

Suzuki, M.T., Giovannoni, S.J. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Applied and Environmental Microbiology. 62(2), 625-630.

Takechi, M., Kawamura, S. 2005. Temporal and spatial changes in the expression pattern of multiple red and green subtype opsin genes during zebrafish development. Journal of Experimental Biology. 208, 1337-1345.

Tamura, K., Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitrochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution. 10, 512-526.

Tamura, K., Dudley, J., Nei, M., Kumar, S. 2007. MEGA4: Molecular Evolution Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution. 10.1093/molbev/msm092.

Taylor, J.S., Braasch, I., Frickey, T., Meyer, A., Van de Peer, Y. 2003. Genome duplication, a trait shared by 22,000 species of ray-fined fish. Genome Research. 13, 382-390.

Temple, S.E., Plate, E.M., Ramsden, S., Haimberger, T.J., Roth, W-M. 2006. Seasonal cycle in vitamin A1/A2-based visual pigment composition during the life history of coho salmon (*Oncorhychus kisutch*). Journal of Comparative Physiology A: Sensory, Neural, and Behavioural Physiology. 192, 301-313.

Terai, Y., Mayer, W.E., Klein, J., Tichy, H., Okada, N. 2002. The effect of selection on a long wavelength-sensitive (LWS) opsin gene of Lake Victoria cichlid fishes. Proceedings of the National Academy of Science of the United States of America. 99, 15501-15506.

Terai, Y., Seehausen, O., Sasaki, T., Takakashi, K., Mizoiri, S., Sugawara, T., Sato, T., Watanabe, M., Konijnendijk, N., Mrosso, H.D.J., Tachida, H., Imai, H., Shichida, Y., Okada, N. 2006. Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. Public Library of Science Biology. 4, 2244-2251.

Thompson, J.D., Higgins, D.G., Gibson, T.J. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. Nucleic Acids Research. 22, 4673-4780.

Thompson, J.R., Marcelino, L.A., Polz, M.F. 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR.' Nucleic Acids Research. 30(9), 2083-2088.

van Noort, V., Worning, P., Ussery, D.W., Roscher, W.A., Sinder, R.R. 2003. Strand misalignments lead to quasipalindrome correction. Trends in Genetics. 19, 365-369.

Van de Peer, Y., De Wachter, R., 1994. TREECON for Windows: a software package for the construction and drawing of evolutionary tree for the Microsoft Windows environment. Computer Applications in the Biosciences. 10, 569-570.

Verrelli, B.C., Tishkoff, S.A. 2004. Signatures of selection and gene conversion associated with human colour vision variation. American Journal of Human Genetics. 75, 363-375.

Volkov, A.A., Shao, Z., Arnold, F.H. 1999. Recombination and chimeragenesis by *in vitro* heteroduplex formation and *in vivo* repair. Nucleic Acids Research. 27(18), i-vi.

Wagner, A., Blackstone, N., Cartwright, P., Dick, M., Misof, B., Snow, P., Wagner, G.P., Bartels, J., Murtha, M., Pendleton, J. 1994. Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. Systematic Biology. 43(2), 250-261.

Ward, M., Churcher, A.M., Dick, K.J., Laver, C.R.J., Owens, G.L., Polack, M.D., Ward, P.R., Breden, F., Taylor, J.S. 2008. Four Long Wave-Sensitive (LWS) opsins in guppies (Poecilia reticulata) are defined by amino acid substitutions at key functional sites. BMC Evolutionary Biology. In revision.

Weadick, C.J., Chang, B.S.W. 2007. Long-wavelength sensitive visual pigments of the guppy (*Poecilia reticulata*): six opsins expressed in a single individual. BMC Evolutionary Biology. 7 (supplement 1), S1-S11.

Weitz, C.J., Miyake, Y., Shinzato, K., Montag, E., Zrenner, E., Went, L.N., Nathans, J. 1992. Human tritanopia associated with two amino acid substitutions in the blue-sensitive opsin. American Journal of Human Genetics. 50, 498-507.

White, E.M., Church, S.C., Willoughby, L.J., Hudson, S.J., Partridge, J.C. 2005. Spectral irradiance and foraging efficiency in guppy, *Poecilia reticulata*. Animal Behavior. 69(3), 519-527.

Whitemore, A.V., Bowmaker, J.K. 1989. Seasonal variation in cone sensitivity and short-wave absorbing visual pigments in the rudd *Scardinius erthrophthalmus*. Journal of Comparative Physiology A. 166, 103-115.

Winderickx, J., Battisti, L., Hibiya, Y., Motulski, A.G., Deeb, S.S. 1993. Haplotype diversity in the human red and green opsin genes: evidence for frequent sequence exchange in exon 3. Human Molecular Genetics. 2(9), 1413-1421.

Wolf, S., Sharpe, L.T., Hans-Jűrgen, A.S., Knau, H., Weitz, S., Kioschis, P., Poustka, A., Zrenner, E., Lichter, P., Wissinger, B. 1999. Direct visual resolution of gene copy number in the human photopigment gene array. Investigative Ophthalmology and Visual Research. 40(7), 1585-1589.

Yokoyama, S., Radlwimmer, F.B. 1999. The molecular genetics of red and green colour vision in mammals. Genetics. 153, 919-932.

Yokoyama, S., Radlwimmer, F.B. 2001. The molecular genetics and evolution of red and green colour vision in vertebrates. Genetics. 158, 1697-1710.

Zylstra, P., Rothenfluh, H.S., Weiller, G.F., Blanden, R.V., Steele, E.J. 1998. PCR amplification of murine immunoglobulin germline V genes: strategies for minimization of recombination artefacts. Immunology and Cell Biology. 76, 395-405.