Oppon gene duplication and diversification in the guppy, a model for sexual selection

Margarete Hoffmann1, Namita Tripathi1, Stefan R. Henz1, Anna K. Lindholm2, Detlef Weigel1, Felix Breden3,† and Christine Dreyer1,8,†

1Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen D72076, Germany
2School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia
3Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 156, Canada

Identification of genes that control variation in adaptive characters is a prerequisite for understanding the processes that drive sexual and natural selection. Male coloration and female colour perception play important roles in mate choice in the guppy (Poecilia reticulata), a model organism for studies of natural and sexual selection. We examined a potential source for the known variation in colour perception, by analysing genomic and complementary DNA sequences of genes that code for visual pigment proteins. We find high sequence variability, both within and between populations, and expanded copy number for long-wave sensitive (LWS) opsin genes. Alleles with non-synonymous changes that suggest dissimilar spectral tuning properties occur in the same population and even in the same individual, and the high frequency of non-synonymous substitutions argues for diversifying selection acting on these proteins. Therefore, variability in tuning amino acids is partitioned within individuals and populations of the guppy, in contrast to variability for LWS at higher taxonomic levels in cichlids, a second model system for differentiation owing to sexual selection. Since opsin variability parallels the extreme male colour polymorphism within guppy populations, we suggest that mate choice has been a major factor driving the coevolution of opsins and male ornaments in this species.

Keywords: guppy; LWS opsin; gene duplication; gene diversification; visual pigment; sexual selection

Abbreviations: LWS; long-wave sensitive; SWS; short-wave sensitive; RH1; rhodopsin; RH2; rhodopsin-like 2; cDNA; complementary DNA; TM; transmembrane (domain); ID; intradisc (domain); SNP; single nucleotide polymorphism; EST; expressed sequence tag

1. INTRODUCTION

The guppy (Poecilia reticulata) is one of the premier model organisms for the study of phenotypic variation owing to divergent natural selection and sexual selection (Magurran & Ramnarine 2005). Male guppy colour patterns in particular are highly polymorphic, both within and between populations. This variation is a consequence of different evolutionary and environmental contexts, including predators (Endler 1980), carotenoid availability (Grether et al. 1999) and female preferences (Houde 1997). Although females overall prefer conspicuous males (Endler & Houde 1995), female preference also varies within (Brooks & Endler 2001) and between populations (Breden & Stoner 1987; Houde & Endler 1990; Endler & Houde 1995). Variation in sensitivity of cone opsins has been suggested as a mechanism explaining this variation in female preference (Endler 1992). Despite the importance of visually mediated sexual selection in the evolution of the mating system of the guppy, very little is known about the diversification of opsin proteins, which help tune absorbance of visual pigments in vertebrates and thus play an essential role in colour discrimination (Yokoyama & Yokoyama 1996).

Teleost fishes express four different cone opsins, which determine sensitivity in red and orange (LWS), green (rhodopsin-like (RH2) or KFH-G), blue and violet (short-wave sensitive 2 (SWS2)) and UV (short-wave sensitive 1 (SWS1)) regions of the spectrum; these cone types also correspond to phylogenetically distinct gene families (Yokoyama 2000; Trezise & Collin 2005). The five sites rule suggests that replacement at a few tuning sites can explain major shifts in maximum absorption of vertebrate cone opsins (Yokoyama & Radlwimmer 1998; Kochendoerfer et al. 1999). Guppy retinal absorption spectra indicate sharp maxima for absorption of UV at 359 nm, violet/blue at 408 nm and green light at 464 nm, with a broader peak of absorbed orange and red light between 520 and 580 nm. This latter peak could be resolved into three maxima at approximately 533, 548 and 572 nm, and interestingly, individuals differ in the occurrence of these three absorption maxima, suggesting allelic variation in LWS opsin genes (Archer et al. 1987; Archer & Lythgoe 1990; Endler et al. 2001).

In the zebrafish, Danio rerio (Chinen et al. 2003), and in the medaka, Oryzias latipes (Matsumoto et al. 2006), LWS opsins are present in two tandemly duplicated copies. Two isoforms have also been reported in Lucania goodei that

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2006.3707 or via http://www.journals.royalsoc.ac.uk.

* Author for correspondence (christine.dreyer@tuebingen.mpg.de).
† F.B. and C.D. contributed equally to the design of the guppy opsin project.
2. MATERIAL AND METHODS

We examined the laboratory strain Wild Istanbul, as well as offspring from wild-caught guppies that had been kept in the laboratory for multiple generations (table 2).

(b) cDNA library and primer design

(i) Oropuche-2 retina library

Retinas were dissected from five adult females from an Oropuche River, Trinidad, stock population and frozen in liquid nitrogen. Total RNA was extracted using a Qiagen RNAeasy Miniprep kit. After reverse transcription, complementary DNA (cDNA) was cloned into pDNRlib vector using a creatorSMART cDNA construction kit (Clonetech). The cDNA was amplified by 18 cycles of long-distance PCR using BD Advantage 2 PCR enzyme mixture, following the manufacturer’s protocols.

(ii) Quare-6 and Tranquille embryo libraries

Late- and very-late-eyed embryos were dissected from pregnant females of Quare-6 and Tranquille stock populations. Total RNA was extracted using a Qiagen RNAeasy Midiprep kit and enriched for polyA+ RNA using Qiagen Oligotron on a spin column. cDNA was cloned as described above, except that the primer extension protocol included only 6–8 cycles of PCR amplification. As part of an expressed sequence tag (EST) sequencing project, cloned cDNAs were isolated using Qiagen MagAttract minipreps and sequenced from the 5' end or both ends using pDNRib-specific vector primers on a Hitachi ABI 3730XL DNA Analyzer. cDNA sequences were identified by NCBI BLASTN and BLASTX (C. Dreyer, M. Hoffmann, C. Lanz, M. Riester, S. R. Henz, and D. Weigel, manuscript in preparation).

(c) PCR analysis of genomic sequences

Consensus primers LWSfw1A (5'-TCT TAT CAG TCT TCA CCA ACG G-3') and LWSrev5 (5'-CAT GAC TAT AAC CAT CCT GG-3') were designed from the alignment of Pseudochromis reticulata LWS cDNA with LWS sequences from medaka (O. latipes, AB001604), blue fin killifish (L. goodei, AF296740) and three cichlids (Dimidiocromis compressiceps, AF247131; Astatotilapia velifer, AB090427; and Gaurochromis simpsoni, AB090425). This primer pair amplifies a region spanning part of exon 2, the highly variable exons 3 and 4 (Terai et al. 2002) and part of exon 5. LWSfw1B (5'-TTT TAT TAT CAG TCT TCA CCA ACG GC-3') is a guppy-specific primer designed from the LWS cDNA sequences. Additional LWS haplotype-specific sequencing primers were LWSE-nInsrev (5'-GAA GAA ATG AGG ATG CAG CA-3'), LWS675AG (5'-GTA GCA CAA GAT GAT GAT ACC TT-3'), LWS675GG (5'-GTA CCA CAA GAT GAT GAT ACC GC-3') and LWS675GC (5'-CAT GCA CAA GAT GAT GAT AG-3'). Primers for other opsins loci included: rhodopsin (RH1) (RH1fw2, 5'-AAG TCA GCC TCC TCA TAC-3'), RH1rev, 5'-GCC TGG ATC AAG TGC TCT TAC-3'), RH2-1 (RH2-1fw, 5'-GGA CGA GAC CTC AGT TTC AG-3'), RH2-1rev, 5'-CTC CAG TAA AAC ATG TTC CCT C-3'), and SWS1 (SWS1fw, 5'-TCA CCT TTC GTA CAA GAT CAT CTC-3'); SWS1rev, 5'-AAG AAC GCA GGA GTG ATG AG-3').

Genomic DNA was extracted using a DEnasy Tissue kit (Qiagen), and amplified using Taq polymerase from New England Biolabs and following the protocol: denaturation at 94°C for 5 min, followed by 35 cycles of (94°C, 30 s; 58°C, 30 s; 68°C, 90 s), and a final elongation step of 68°C for 6 min. Sizes of amplification products are given in table 1. Direct sequencing of PCR products for LWS showed that many sites were polymorphic, and therefore PCR products were purified with BioInol Quick Clean and cloned into pGEMTeasy (Promega) vector. T7 and SP6 promoter

Table 1. Variability in guppy opsin cDNAs and genomic fragments.

<table>
<thead>
<tr>
<th>opsin</th>
<th>cDNAs clones among 18 000 ESTs</th>
<th>length coding sequence (bp)</th>
<th>synonymous SNPs</th>
<th>non-synonymous SNPs</th>
<th>length (bp)</th>
<th>synonymous and non-coding SNPs</th>
<th>non-synonymous SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>RH1</td>
<td>148</td>
<td>1062</td>
<td>3</td>
<td>1</td>
<td>497</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>RH2-1</td>
<td>14</td>
<td>1056</td>
<td>0</td>
<td>0</td>
<td>450</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>RH2-2</td>
<td>9</td>
<td>1036</td>
<td>1</td>
<td>0</td>
<td>912</td>
<td>5 ± 1 indel</td>
<td>0</td>
</tr>
<tr>
<td>SWS1</td>
<td>10</td>
<td>1008</td>
<td>0 (6)</td>
<td>0</td>
<td>n.d.</td>
<td>n.d.</td>
<td>0</td>
</tr>
<tr>
<td>SWS2B</td>
<td>2</td>
<td>1056</td>
<td>0 (2)</td>
<td>0</td>
<td>920</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>LWS</td>
<td>5</td>
<td>1074</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Not determined.

** SNPs that could not be confirmed by sequencing of PCR amplification products are given in parentheses.

mid-probably also represent two gene copies (Fulcher et al. 2003). Phylogenetic analysis suggests that these duplications are independent (Chinen et al. 2003). In contrast, Southern blot analysis indicated that there is only a single LWS opsin copy in the cichlid Neochromis nigricans (Terai et al. 2002). While a given cichlid species generally has only one or two LWS isoforms, an amazing degree of interspecific LWS sequence variation has been documented (Terai et al. 2002; Carleton et al. 2005; Parry et al. 2005; Spady et al. 2005). Since cichlid species in African rift lakes have only arisen very recently, variation must have evolved within a comparatively short time frame, indicating diversifying selection acting on visual perception.

The molecular genetic basis underlying sexually selected characters can reveal the pattern of selection that has been acting on these characters and help to discriminate among competing models explaining diversification and speciation owing to sexual and natural selection. Therefore, we compared LWS opsin sequences from several individuals and populations across the geographical range of the guppy. We focused on LWS polymorphisms, because these may provide a molecular basis for the broad and variable red sensitivity of the guppy retina, and because the perception of red and orange signals plays a major role in mate choice of guppies.
primers were used to sequence clones. One PCR product, amplified with primers fw1A and rev5 from a Xiphophorus primer, was sequenced. One PCR product, (ID 1–6) are shown above. Amino acid position follows the bovine RH1 reference. Amino acids previously described as critical variable sites represent characteristic differences between green- and red-sensitive vertebrate pigments (Kochendoerfer et al. 1999). In addition, amino acids that differ between guppy OR6-3 and OR6-4 are colour-coded yellow and blue across all LWS sequences. Cichlids I and II represent the major allelic types from Lakes Victoria and Nabugabo (Terai et al. 2002). For comparison, the last line shows residues of guppy green opsin RH2-1. Wavelength of maximal absorption of proteins reconstituted with 11-cis-retinal in vitro (λmax) is given in parenthesis, if published. Accession numbers are given in table 3.

**Figure 1.** Amino acid substitutions in LWS opsin protein sequences. Transmembrane domains (TM 1–7) and intradisc domains (ID 1–6) are shown above. Amino acid position follows the bovine RH1 reference. Amino acids previously described as critical variable sites represent characteristic differences between green- and red-sensitive vertebrate pigments (Kochendoerfer et al. 1999). In addition, amino acids that differ between guppy OR6-3 and OR6-4 are colour-coded yellow and blue across all LWS sequences. Cichlids I and II represent the major allelic types from Lakes Victoria and Nabugabo (Terai et al. 2002). For comparison, the last line shows residues of guppy green opsin RH2-1. Wavelength of maximal absorption of proteins reconstituted with 11-cis-retinal in vitro (λmax) is given in parenthesis, if published. Accession numbers are given in table 3.

<table>
<thead>
<tr>
<th>domains</th>
<th>ID1</th>
<th>TM1</th>
<th>TM3</th>
<th>TM4</th>
<th>ID3</th>
<th>TM5</th>
<th>TM6</th>
<th>TM7</th>
</tr>
</thead>
<tbody>
<tr>
<td>position</td>
<td>0 0 0 0 1 1 1 1</td>
<td>1 1 1 2 2 2 2 2</td>
<td>2 2 2 2 3</td>
<td>3 5 5 6 6 3 3 4</td>
<td>4 7 8 8 9 1 3 3</td>
<td>4 7 7 7 8 2</td>
<td>7 5 6 5 8 1 4 0</td>
<td>1 0 2 4 9 0 3 7</td>
</tr>
<tr>
<td>cichlids I</td>
<td>R L A V V V Y S A A A A A V Y Y I L T T V Y C T I</td>
<td>cichlids II</td>
<td>R L S I V Y A A G S F F F F F F F F M Y I A V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR6-3 (variant 1)</td>
<td>K Y S V Y V S A A A V Y Y I A L I Y C T V</td>
<td>OR6-4 (variant 2)</td>
<td>R L S V V F S A G S A F Y Y I G H I Y C T V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RH2-1</td>
<td>R L L C T F Y T A A A A B Y I F L F L A V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(d) **DNA blot analysis**

Genomic DNA was isolated from tail muscles of at least six fishes from each of Quare, Tranquile and Cumaná strains, and pooled for each strain. DNA was extracted using QIAGEN genomic-tip 500/G columns. Six micrograms of digested DNA was concentrated by ethanol precipitation, reconstituted with 11-cis-retinal and transfected to a positively charged nylon membrane (Roche). DNA was UV cross-linked (STRATALINKER 2400) and pre-hybridized in 5× SSC, 0.02% (w/v) SDS, 0.1% (w/v) N-lauroylsarcosine and 1% blocking reagent overnight at 62°C. Cloned RH1, RH2-1, SWS1 and LWS fragments were labelled using the PCR DIG-Probe Synthesis Kit (Roche). Membranes were hybridized overnight at 60°C. The membranes were washed three times in 0.1% SDS, 2× SSC (30 min, 60°C), twice in 0.1% SDS, 2× SSC (20 min, 62°C), and twice in 0.1% SDS, 0.5× SSC (15 min, 62°C). Membranes were treated with blocking solution, followed by 1:10 000 (v/v) anti-DIG-antibodies (Roche). Detection was performed using CSPD chémiluminescent substrate (Roche). Blots were exposed to Lumi-Film Chemiluminescent Detection Film (Roche) for 10–60 min.

(e) **Comparison with other vertebrate long-wave sensitive sequences**

GCG (Wisconsin Package v. 10.3, Accelrys, Inc., San Diego, CA) or CLUSTALW (v. 1.81, Thompson et al. 1997) was used to align guppy cDNAs with LWS sequences from GenBank from other species of fish, chicken and human. Our alignment follows the numbering system of Yokoyama & Radlwimmer (1998) to facilitate comparison to substitutions known to affect maximal absorbance of other vertebrate visual pigments.

Given the high variability in these sequences and the possibility that some of this variation were owing to PCR artefacts, a conservative approach was taken for reporting single nucleotide polymorphisms (SNPs). Only SNPs that occurred in at least two individuals were analysed. In addition, four sequences with frame shifts were omitted from the analyses. Sequences were aligned using MUSCLE (Edgar 2004) and CLUSTALX (Thompson et al. 1997), and the program COLLABE (v. 1.2, http://darwin.uvigo.es/software/collaps.html) was used to identify unique haplotypes in the genomic sequences. A BIONJ tree (Gascuel 1997) was constructed using the X. helleri sequence as outgroup and using SPLITSTREE4 software (Huson & Bryant 2006). An alignment of the set of unique haplotypes was submitted as a population study to GenBank.

(f) **Molecular evolution analysis**

The rate of non-synonymous nucleotide substitutions per non-synonymous site, relative to the rate of synonymous nucleotide substitutions per synonymous site (Ka/Ks), across all amino acid sites was calculated for the partial LWS genomic sequences and the corresponding part of the LWS cDNAs. Branch-specific ratios were estimated in at least two individuals were analysed. In addition, four sequences with frame shifts were omitted from the analyses. Sequences were aligned using MUSCLE (Edgar 2004) and CLUSTALX (Thompson et al. 1997), and the program COLLABE (v. 1.2, http://darwin.uvigo.es/software/collaps.html) was used to identify unique haplotypes in the genomic sequences. A BIONJ tree (Gascuel 1997) was constructed using the X. helleri sequence as outgroup and using SPLITSTREE4 software (Huson & Bryant 2006). An alignment of the set of unique haplotypes was submitted as a population study to GenBank.

3. RESULTS

(a) **Diversity of opsin cDNAs**

We isolated cDNAs of RH1 and five different cone opsins from libraries of whole embryos and adult retinas of the
guppy (table 1) in the course of an EST project, whose details will be reported elsewhere. The abundance of the different cDNAs in the EST libraries gives a first hint of their relative expression levels. RH1, for which we found relatively little sequence variability, was the most abundant opsin (table 1). The RH1 cDNAs from the Quare strain were polymorphic at three sites, of which one was non-synonymous, and fell into two haplotypes. cDNAs from the other strains were polymorphic at a single synonymous site each.

Alignments of the five different cone opsins, SWS1 (UV-sensitive), SWS2B (violet-sensitive), RH2-1 and RH2-2 (green-sensitive) and LWS (red-sensitive) are shown in the electronic supplementary material, figure 1. Two different groups of RH2 cDNAs were isolated, similar to the two different RH2 forms in other species (Neafsey & Hartl 2005; Parry et al. 2005; Matsumoto et al. 2006). Guppy RH2-1 and RH2-2 are distinguished by at least two previously described candidates for tuning amino acid substitutions: E139Q and S312A, corresponding to positions 122 and 292 in bovine RH1 (22). The overall amino acid identity between RH2-1 and RH2-2 is approximately 80%.

The five full-length LWS opsin cDNAs, all from the Oropuche population, cluster into two types: LWS_OR6-3 and LWS_OR6-4. These differ at nine sites, all of which are non-synonymous (figure 1 and electronic supplementary material, figure 1). Of these, A180S in the fourth transmembrane (TM) domain was labelled in red and green because it is one of the key tuning amino acids (Yokoyama & Radlwimmer 1998; Kochendoerfer et al. 1999; Yokoyama 2000). There were 26 polymorphic sites in this region, with many individuals being heterozygous at multiple sites (electronic supplementary material, figure 2). The SNP density was at least three times higher in LWS than in RH1 and at least five times higher than in the other cone opsins (table 1).

The genomic LWS product appeared to be homozygous at all sites in a single Quare individual, LQu47 f. Two individuals from the Aripo and Tacarigua populations shared the same exon sequence, corresponding to the sequence of LWS_OR6-4 cDNA, but were heterozygous in introns. That this set of SNPs in the exons was observed in three out of 11 individuals and one cDNA, despite high population heterozygosity at multiple sites, suggests that these non-synonymous SNPs are in strong linkage disequilibrium. PCR amplification with allele-specific primers after digestion of DNA further supported the linkage of these SNPs, as explained in electronic supplementary material, figure 2.

All other individuals were heterozygous at non-synonymous sites and thus had the potential to express multiple LWS proteins. Three Quare, one Oropuche and one Tunapuna individual were heterozygous in at least 10 positions, seven of which caused amino acid replacements (electronic supplementary material, figure 2). Cumana and PV6 individuals were polymorphic for an indel, with additional SNPs at up to 17 positions, six of which caused non-synonymous substitutions. Two Tranquille individuals had 18 heterozygous positions; these comprise the SNPs found in the eastern strains plus the ones found in individuals from the Aripo and Tacarigua populations.

(b) Long-wave sensitive opsin copy number
We determined the minimum number of LWS opsin gene copies by genomic DNA blots. Using restriction enzymes that do not cut within the cDNA sequences, single fragments were detected for RH1, RH2-1 and SWS1 (figure 2a). In contrast, multiple LWS fragments were seen (figure 2a,b), suggesting a minimum of two gene copies, given that some of these different bands could be owing to allelic variation.

The two LWS opsin cDNA sequences differ in the presence of a PvuII restriction site, allowing for their differential detection. The observed hybridization pattern indicates that there are additional PvuII sites flanking the gene loci, and it favours the presence of a minimum of two copies in tandem in all three strains analysed (figure 2c). The pattern was the same after double digestes with PvuII and either BamHI or HindIII (data not shown). Different individuals from the Cumana and Quare strains had the same fragment pattern, in agreement with results on pooled DNA from individuals of the same strains (data not shown). However, restriction length polymorphisms were observed between strains and also within the Tranquille strain (figure 2c).

(e) Genomic diversity of long-wave sensitive opsin genes
We further investigated variation in LWS opsin genes by direct sequencing of PCR products from 16 individuals representing nine populations (table 2 and electronic supplementary material, figure 2). The examined 920 bp fragment codes for most of the tuning amino acids (Yokoyama & Radlwimmer 1998; Kochendoerfer et al. 1999; Yokoyama 2000). There were 26 polymorphic sites in this region, with many individuals being heterozygous at multiple sites (electronic supplementary material, figure 2). The SNP density was at least three times higher in LWS than in RH1 and at least five times higher than in the other cone opsins (table 1).

To determine the haplotypes and their distribution between populations, we sequenced 193 clones from the 16 individuals investigated above (table 2). The PCR products from all individuals but one were variable, and all SNPs identified by direct sequencing of genomic amplification products were confirmed with cloned products. The simultaneous amplification of several copies of a gene will probably give rise to some recombinant products owing to template switching (Odelberg et al. 1995). Therefore, we restricted our analysis to 76 clones that were observed at least twice, resulting in 15 unique haplotypes (figure 3). Of these 76 sequences, 20 and 50% encoded protein variants 1 and 2, respectively, corresponding to cDNA sequences OR6-3 and OR6-4 (figure 3) table 3.

(e) Molecular evolution of long-wave sensitive haplotypes
We used the BIONJ algorithm (Gascuel 1997) to generate a neighbour-joining tree of the 15 unique haplotypes listed in figure 3. The tree has three major branches (figure 4). As an outgroup we used genomic sequences from a swordtail (X. hellerti) individual, which was polymorphic at only two synonymous positions. The position of the outgroup suggests that the deepest split is between variant 1 and a group comprising two other variants. Variants 1
Figure 2. Genomic DNA blots. (a) Genomic DNA from females (F) and males (M) of the Wild Istanbul strain digested with the indicated enzymes and hybridized with probes for opsins RH1, RH2-1, LWS and SWS1. A BamHI site in the RH2-1 cDNA sequence results in two RH2-1 opsin bands. (b) Pooled DNA of guppies from the Tranquille (Tr), Cumana (C) and Quare (Qu) strains digested with HindIII and hybridized with a LWS probe. (c) DNA from individual females and males of indicated strains digested with PvuII and hybridized with a LWS probe.

Table 2. Number of LWS proteins and haplotypes in 16 individuals.

<table>
<thead>
<tr>
<th>individual</th>
<th>strain</th>
<th>population</th>
<th>proteins</th>
<th>haplotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>LQu47 f</td>
<td>Lower Quare</td>
<td>Quare R., Trinidad</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>LQu49 m</td>
<td>Lower Quare</td>
<td>Quare R., Trinidad</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Qu135 f</td>
<td>Quare 206-5</td>
<td>Quare R., Trinidad</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Qu196 f</td>
<td>Quare 206-5</td>
<td>Quare R., Trinidad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Or56 m</td>
<td>Oro 201-5</td>
<td>Oro 206-5, Trinidad</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Tu128 m</td>
<td>Tunapuna</td>
<td>Tunapuna, Trinidad</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Ta16 m</td>
<td>Tacariqua</td>
<td>Tacariqua, Trinidad</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Tr130 m</td>
<td>Tranquille</td>
<td>Tranquille, Trinidad</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Tr131 f</td>
<td>Tranquille</td>
<td>Tranquille, Trinidad</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>AR138 m</td>
<td>ARDC</td>
<td>Aripo R., Trinidad</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ari18 f</td>
<td>Lower Aripo</td>
<td>Aripo R., Trinidad</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>En25 f</td>
<td>CCFR</td>
<td>Cumaná, Venezuela b</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>En28 m</td>
<td>CCFR</td>
<td>Cumaná, Venezuela</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>En157 f</td>
<td>CCFAH</td>
<td>Cumaná, Venezuela</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>En158 f</td>
<td>CCFNX</td>
<td>Cumaná, Venezuela</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>PV6107 m</td>
<td>PV6FIC</td>
<td>Rio San Miguel, Venezuela</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Xiphophorus helleri</td>
<td>pet store</td>
<td>unknown</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

a Not determined.
b Known in the aquarium trade as Endler’s Livebearer (Alexander & Breden 2004).
and 2 have further diversified in a way that suggests the existence of seven protein isoforms (figure 3) occurring in three phylogenetically distinct clades (figure 4). Variant 1 was found across the range of the guppy, in both the distinct eastern and western Trinidadian strains (Fajen & Breden 1992) as well as in the Venezuelan populations. In contrast, variant 2 was found only in Trinidad. Variant 3 in the Venezuelan populations Cumana and PV6 apparently replaces it.

The rate of non-synonymous nucleotide substitutions per non-synonymous site, relative to the rate of synonymous nucleotide substitutions per synonymous site (Ka/Ks) can reveal whether selection is likely to have acted on a set of sequences. Low Ka/Ks ratios are indicative of purifying selection, while high Ka/Ks ratios are taken as an indication of diversifying selection. We used DNASP (v. 4.10.4; Rozas et al. 2003) to estimate branch-specific Ka/Ks ratios on the LWS phylogeny (figure 5). The highest Ka/Ks values were found to coincide with polymorphisms occurring in the fourth TM domain.

(f) Functional implications of long-wave sensitive diversity in the guppy

According to the three sites rule (Yokoyama & Yokoyama 1996; Yokoyama & Radlwimmer 1998) for tuning of LWS opsins, polarity changes caused by the substitutions S180A, Y277F and T285A are the most important for tuning vertebrate LWS (red-sensitive) opsins to shorter, green-shifted wavelengths. Of these three amino acids, only S180A is polymorphic between the LWS variants 1 and 2. Two additional polymorphisms, F131Y and F194Y, also cause amino acid polarity changes. All of the intraspecific substitutions found coincide with polymorphisms in LWS opsins of other fishes (figure 1). In particular, LWS opsins of different species of cichlids from Lakes Victoria and Nabugabo are polymorphic for at least five of the positions that are variable in guppies, and the cichlid LWS opsins have similarly been grouped into two classes based on these substitutions (Terai et al. 2002).

To contrast similarities owing to descent and those owing to convergent positive selection, we applied the BIONJ algorithm both to the entire guppy sequences, and only to the amino acids shown in figure 1, several of which have been implicated as having tuning function. The tree based on complete sequences shows clustering of the different variants from the same species, including guppies, L. goodei, medaka and zebrafish, as well as clustering of the sequences from two cichlid species (figure 6a). As expected, LWS opsins of cichlids and guppies are most closely related to those of medaka and fugu and most distant from zebrafish. This molecular phylogeny suggests that the LWS opsin variants found in different taxa have arisen by independent duplications in each lineage. In contrast to the phylogenetically consistent tree of the entire sequences, analysis of only the discriminating amino acids shown in figure 1 suggests functional similarity of guppy OR6-4 and cichlid variant II, while OR6-3 clusters with both L. goodei variants (figure 6b).
pattern suggests convergent evolution of functionally distinct LWS opsins in different fish taxa.

4. DISCUSSION

We found that the cone opsins of guppies are highly polymorphic, both within and between populations, resulting in at least two different green-sensitive opsins and two highly differentiated LWS opsin isoforms. While RH1, RH2-1 and SWS1 appear to be single copy genes, LWS is found in at least two copies per individual. The 15 different LWS opsin haplotypes identified from nine strains can encode seven different proteins, which can clearly be grouped into three distinct forms. Only variant 1 was found in populations from both Trinidad and Venezuela, while variants 2 and 3 were restricted to Trinidad and Venezuela, respectively (figures 3 and 4). The prevalence of non-synonymous substitutions known to change maximum absorbance of visual pigments, along with the high ratio of non-synonymous to synonymous substitutions, suggests strong diversifying selection of these proteins, especially in the functionally important TM domain 4 (figure 5).

Our finding of five different cone opsins is compatible with microspectrophotometric (MSP) studies of the guppy retina, which suggest the presence of UV, blue/violet, green and variable red/orange-sensitive cone opsins. These long wavelength-sensitive cones are by far the most polymorphic class, in which three different absorption peaks at 533, 548 and 572 nm can be distinguished (Archer et al. 1987; Archer & Lythgoe 1990). Although we found two different RH2 isoforms, MSP data did not indicate the presence of functionally different green sensitive cones, suggesting that the RH2 differences are too subtle to be detected. Alternatively, expression of some visual pigments could be developmentally regulated, as has been found for green opsins in cichlids (Spady et al. 2006).

Archer & Lythgoe (1990) suggested the existence of two LWS opsins with distinct absorption maxima, with co-expression of both causing a third absorption peak, to account for individuals with a single, two or three absorption peaks for long wavelength light. Our data provide a possible alternative explanation, since at least some individuals have three different LWS cones, most of the other individuals having either one or another. Our genomic data are in broad agreement with this conclusion, as most individuals have a repertoire of LWS genes that should allow for at least two different proteins. While the maximum number of isoforms observed within one individual was four, a quarter of individuals seemed to have only one type of LWS protein (table 2). There was only partial overlap between the populations studied by
Archer & Lythgoe (1990) and the strains included in our survey, and the percentage of individuals with a single LWS opsin could differ between populations.

Our estimates of LWS haplotype number are conservative, although a low frequency of artefacts owing to template switching during PCR cannot be excluded. Several factors may have resulted in an underestimate of the number of haplotypes per individual. For example, we could not directly select for LWS alleles represented by cDNA OR6-3 (variant 1) with allele-specific primers. In addition, point mutations in the regions covered by the consensus primers might have prevented detection of some isoforms.

Archer & Lythgoe (1990) and the strains included in our survey, and the percentage of individuals with a single LWS opsin could differ between populations.

Our estimates of LWS haplotype number are conservative, although a low frequency of artefacts owing to template switching during PCR cannot be excluded. Several factors may have resulted in an underestimate of the number of haplotypes per individual. For example, we could not directly select for LWS alleles represented by cDNA OR6-3 (variant 1) with allele-specific primers. In addition, point mutations in the regions covered by the consensus primers might have prevented detection of some isoforms.

Long-wave sensitive opsin genes have been compared between cichlid species that have only recently evolved in East African rift lakes (Terai et al. 2002; Carleton et al. 2005; Parry et al. 2005; Spady et al. 2005). Although DNA blot experiments did not indicate multiple LWS copies, genomic PCR analysis revealed 14 different LWS alleles in cichlids from Lakes Victoria and Nabugabo, with two alleles found in most species (Terai et al. 2002). In cichlids, differentiation of colour patterns and of the visual system...
have been associated with adaptation to different habitats (Parry et al. 2005), and both of these factors have been proposed as essential components of speciation.

The situation in cichlids, in which allelic diversity is found predominantly between species (Terai et al. 2002), contrasts with the one in guppies, in which eight strains of the same species contain at least 15 different haplotypes encoding seven different proteins, with obvious geographical differentiation. Importantly, much of the variation in LWS opsins occurred within populations and even within individuals in guppies. Guppies can live in different photic environments that are distinguished by degree of transparency and amount of canopy (Endler 1991, 1995), but the main forces driving differentiation of male colour patterns are thought to be sexual selection (Houde 1997) and predation (Olendorf et al. 2006). Further detailed analysis of a greater number of specimens from upper and lower ranges of several northern Trinidad river systems might reveal whether a correlation exists between opsin gene pools and male ornamentation. How this colour variation, one of the highest in vertebrates (Haskins 1961), is maintained continues to be an important question in evolutionary biology (Olendorf et al. 2006). The colour diversity is due to as many as 40 genes with polymorphic alleles for male coloration, most of which are linked to the sex chromosomes (reviewed in Lindholm & Breden 2002).

In contrast to the highly variable patterns of guppies, five distinct male colour morphs coexist in Poecilia parae and female mate choice may counteract loss of the less frequent morphs (Lindholm et al. 2004). In this context, it would be interesting to know to what extent opsin gene diversification has occurred in closely related poeciliids such as P. parae.

Guppies also stand out among species studied for sexual selection in that female preference functions are variable within and between populations (Houde & Endler 1990; Endler & Houde 1993; Brooks & Endler 2001). Sensitivity to UV light and the detection of black and iridescent ornaments play a role in female preferences in guppies (Houde 1997; Smith et al. 2002). However, sensitivity to red and orange wavelength is likely to be particularly important for differentiation because red and orange ornaments are the most consistent targets of female choice (Endler & Houde 1995; Houde 1997; Brooks & Endler 2001). Further, sensitivity to red wavelengths shows heritable variation in guppies (Endler et al. 2001), and changes in wavelength sensitivity are expected to alter female preferences. Hence, differentiation in LWS opsins could select for variation in male nuptial colour patterns (Endler 1992, 1993; Endler et al. 2001). Alternatively, the variation in red and orange ornaments themselves might select for diversification of red colour perception. Either way, the highly polymorphic nature of opsins, and the partitioning of some of this variation between individuals within populations, provides a potential molecular framework for future studies of coevolution between visual perception and male ornamentation.

5. DATA DEPOSITION

The sequences reported in this paper have been deposited at GenBank (accession nos DQ168860, DQ234858–DQ234861, DQ435005–DQ435019, DQ912023–DQ912026).

We thank Christa Lanz and Heike Keller for expert assistance with sequencing, Markus Riester for database construction, Stephen Russell for critical reading of the manuscript and Richard Clark for helpful suggestions. We also thank two anonymous reviewers for their helpful comments. Bernie Crespi, Werner Mayer, Arne Mooers and Patrik Nosil read an earlier draft of this manuscript. Guppies were provided by David Reznick, Anne Maguran, Axel Meyer and Manfred Schartl, or collected by F.B. in Venezuela in collaboration with Donald Taphorn of UNELLEZ (permit no. 0497) and in Trinidad with the permission of the Ministry of Agriculture, Land and Marine Resources. This work was funded by grants from the Natural Sciences and Engineering Research Council of Canada, the German Academic Exchange Service (DAAD), which enabled a sabbatical visit of F.B. at the Max Planck Institute, and the Max Planck Society, of which D.W. is a director.

REFERENCES


