

ORIGINAL ARTICLE

Divergent selection for opsin gene variation in guppy (*Poecilia reticulata*) populations of Trinidad and Tobago

A Tezuka¹, S Kasagi², C van Oosterhout³, M McMullan³, WM Iwasaki¹, D Kasai², M Yamamichi⁴, H Innan⁴, S Kawamura² and M Kawata¹

The guppy is known to exhibit remarkable interindividual variations in spectral sensitivity of middle to long wavelength-sensitive (M/LWS) cone photoreceptor cells. The guppy has four M/LWS-type opsin genes (*LWS-1*, *LWS-2*, *LWS-3* and *LWS-4*) that are considered to be responsible for this sensory variation. However, the allelic variation of the opsin genes, particularly in terms of their absorption spectrum, has not been explored in wild populations. Thus, we examined nucleotide variations in the four M/LWS opsin genes as well as blue-sensitive *SWS2-B* and ultraviolet-sensitive *SWS1* opsin genes for comparison and seven non-opsin nuclear loci as reference genes in 10 guppy populations from various light environments in Trinidad and Tobago. For the first time, we discovered a potential spectral variation (180 Ser/Ala) in *LWS-1* that differed at an amino acid site known to affect the absorption spectra of opsins. Based on a coalescent simulation of the nucleotide variation of the reference genes, we showed that the interpopulation genetic differentiation of two opsin genes was significantly larger than the neutral expectation. Furthermore, this genetic differentiation was significantly related to differences in dissolved oxygen (DO) level, and it was not explained by the spatial distance between populations. The DO levels are correlated with eutrophication that possibly affects the color of aquatic environments. These results suggest that the population diversity of opsin genes is significantly driven by natural selection and that the guppy could adapt to various light environments through color vision changes.

Heredity (2014) **113**, 381–389; doi:10.1038/hdy.2014.35; published online 2 April 2014

INTRODUCTION

Color vision allows organisms to discriminate differences in spectral composition, irrespective of the light intensity (Jacobs and Rowe, 2004), that helps them to find food, avoid predators and select their mates. Visual photoreception is mediated by visual pigments that comprise a protein moiety, opsin and a chromophore, either 11-*cis*-retinal or 11-*cis*-3,4-dehydroretinal (vitamin A₁ or A₂ aldehyde, respectively). Vertebrate visual opsin genes are classified into five types: one type expressed in the rod photoreceptor cells, that is, RH1 ('rhodopsin'), responsible for dim light perception, and four types expressed in the cone photoreceptor cells, that is, middle to long wavelength-sensitive (M/LWS, 'red-green'), RH2 ('green'), SWS2 ('blue') and SWS1 ('violet-ultraviolet') that are responsible for bright light and color perception (Yokoyama, 2000). Mutations in the coding regions of opsin genes can affect the structure of opsin proteins and cause changes in the peak absorption spectra, λ_{\max} , of the chromophore (Yokoyama and Radlwimmer, 1998).

Opsin genes are known to have an important role in ecological adaptation and even in reproductive isolation. In many cichlid fishes, for example, allelic variations of the opsin genes have been reported among populations and closely related species (Terai *et al.*, 2002, 2006). In Lake Victoria, the allelic variation of the M/LWS opsin genes

among populations is explained by their adaptation to different light environments (Terai *et al.*, 2006). Two sympatric cichlids (*Pundamilia pundamilia* and *Pundamilia nyererei*) that inhabit different water depth differ in the allelic compositions of their M/LWS opsin genes with different λ_{\max} , and they are adapted to the spectral composition of their aquatic environment (Maan *et al.*, 2006). This adaptation is also hypothesized to result in a female preference for different male nuptial coloration, where the M/LWS allele is the most sensitive, thereby leading to sympatric speciation (a typical example of sensory-driven speciation) (Seehausen *et al.*, 2008).

The guppy (*Poecilia reticulata*) is an important evolutionary and ecological model organism because of its high phenotypic variation. The male body color patterns are highly polymorphic and are considered to be related to predation pressure and female preference (Houde and Hankes, 1997; Breden, 2006). Females prefer some components of these male color patterns. This female preference is heritable (Houde and Endler, 1990) and varies within and between populations (Endler and Houde, 1995; Houde and Hankes, 1997; Breden, 2006; Schwartz and Hendry, 2007). This suggests that variation in female preference could drive variation in male color patterns (Brooks, 2002), although other selective agents possibly have additional roles (Schwartz and Hendry, 2007). Spectral measurements

¹Department of Ecology and Evolutionary Biology, Graduate School of Science, Tohoku University, Sendai, Japan; ²Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Japan; ³School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, UK and ⁴Department of Evolutionary Studies of Biosystems, Graduate University for Advanced Studies (Sokendai), Hayama, Japan

Correspondence: Professor M Kawata, Department of Ecology and Evolutionary Biology, Graduate School of Sciences, Tohoku University, Sendai 980-8578, Japan.

E-mail: kawata@m.tohoku.ac.jp

or Professor S Kawamura, Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa 277-8562, Japan.

E-mail: kawamura@k.u-tokyo.ac.jp

Received 8 April 2013; revised 3 December 2013; accepted 24 February 2014; published online 2 April 2014

of the cone photoreceptor cells of Trinidadian guppies using microspectrophotometry (MSP) detected a considerable amount of inter-individual variations in the spectral sensitivity of cells to the middle to long wavelength ranges (Archer *et al.*, 1987; Archer and Lythgoe, 1990). These cells were classified into three classes, where λ_{\max} was centered at ~ 533 , 548 and 572 nm, and individuals exhibit either one, two or three of these classes (Archer and Lythgoe, 1990). The MSP absorption curves of the guppy cones were reported to be fitted better by an A₁ chromophore ‘rhodopsin’ template than by an A₂ chromophore ‘porphyropsin’ template that suggests that the spectral variation of the cones is not because of possible differences in A₁ and A₂ chromophore usage among cells (Archer *et al.*, 1987; Archer and Lythgoe, 1990). The 533 and 572 nm cells exhibited a narrow λ_{\max} distribution range, whereas the 548 nm cells had a broad range, implying that the 548 nm cells may represent a group of cells that express both the hypothetical 533 and 572 nm opsin genes simultaneously in the same cell with varying ratios (Archer and Lythgoe, 1990). Endler (1992) suggested that this sensory variation could contribute to differences in female preference.

Hoffmann *et al.* (2007) first reported several different sequences belonging to the M/LWS and other types of opsin genes, and this genetic variation was present both within and between individuals of guppies. Multiple M/LWS sequences were also reported independently by Weadick and Chang (2007). These studies did not demonstrate whether this variation represented alleles, paralogs or both. Subsequently, the paralogous composition of the guppy M/LWS opsins was found to comprise four genes (Ward *et al.*, 2008), that is, *LWS-1*, *LWS-2*, *LWS-3* and *LWS-4*, where the nomenclature of the first three genes follows that in the study by Sandkam *et al.* (2013) and that of the last was renamed in this study, for simplicity. Downstream of two *SWS2* genes (*SWS2-A* and *SWS2-B*), the *LWS-1*, *LWS-2* and *LWS-3* genes are physically linked in this numerical order, with *LWS-3* in the opposite orientation (Watson *et al.*, 2011). *LWS-4* is located in a different linkage group and lacks introns (Ward *et al.*, 2008), except intron 1 (Watson *et al.*, 2010, 2011). Based on the amino acid that corresponds to residue 180 of the human M/LWS opsins that is known to affect λ_{\max} , the *LWS-1*, *LWS-2*, *LWS-3* and *LWS-4* genes were originally designated as *A180*, *P180*, *S180* and *S180r*, respectively. More recently, studies have demonstrated intraretinal variations and age/sex differences in opsin gene expression in guppies (Laver and Taylor, 2011; Rennison *et al.*, 2011). Despite these advances, the genetic basis of interindividual sensory variation and the selection forces that maintain this variation remain unexplored.

In the present study, we investigated nucleotide polymorphisms in four M/LWS opsin genes of guppies in 10 populations from Trinidad and Tobago that inhabited various light environments. In addition, we analyzed the blue-sensitive *SWS2-B* and ultraviolet-sensitive *SWS1* opsin genes for comparison as well as seven non-opsin nuclear loci as (potentially neutral) reference genes. First, we quantified the allelic variation of amino acid sites that are known to affect spectral sensitivity in the M/LWS-type opsin genes, that is, residues 180, 197, 277, 285 and 308 (the numbering follows that of the human M/LWS opsins throughout this paper) (Yokoyama and Radlwimmer, 1998, 2001). Next, we examined the overall pattern of nucleotide variation in the M/LWS opsin genes within and between populations, and determined whether it was consistent with a model of neutral evolution and demographic effects based on comparisons with the reference genes. We detected an allelic variant at residue 180 in *LWS-1*. We also found population genetic evidence for divergent natural selection among populations of nucleotide variation in the

linkage group of *SWS2-B-LWS-3*. We also investigated the relationships between several environmental variables and opsin divergence.

MATERIALS AND METHODS

Sample populations

In Trinidad, guppies were sampled from two rivers (Aripo and Guanapo) within the Caroni Drainage, one river (Yarra) within the Northern Drainage as well as one lake (Pitch Lake) and three rivers (Visigney Beach, Coffee and Vance River) in southern Trinidad (Figure 1). Guppies were sampled from both the upper and lower reaches of Aripo and Guanapo and the lower reaches in Yarra. Samples were also collected from the Lambeau River in Tobago. Adult male and female guppies were collected using seine nets from the shallow parts (<1 m water depth) of rivers and ponds. We randomly sampled individuals from multiple shoals at each sampling site. The fish were killed with an overdose of tricaine methanesulfonate (MS222) and preserved in 99.5% ethanol. A total of 429 individuals were collected from 10 wild populations in these areas of Trinidad and Tobago. The abbreviations of the populations are as follows: Upper Aripo (UA), Lower Aripo (LA), Upper Guanapo (UG), Lower Guanapo (LG), Lower Yarra (LY), Pitch Lake (PL), Visigney Beach (VB), Coffee (CO), Vance River (VR) and Lambeau River (LR) (Figures 1 and 2). The Aripo and Guanapo rivers have a series of waterfalls and weirs. Therefore, migration appears to be biased downstream because of floods, particularly during the wet season (van Oosterhout *et al.*, 2007). Upstream migration over the waterfalls is severely limited, although the rate of gene flow is considerably higher than the mutation rate (Carvalho *et al.*, 1991; van Oosterhout *et al.*, 2006; Barson *et al.*, 2009; Willing *et al.*, 2010). The Caroni drainage is considered to be a source-sink metapopulation, where the downstream populations represent a ‘super sink’ that receives immigrants from the upstream populations (Barson *et al.*, 2009). Northern drainage is isolated from Caroni drainage, but there is disrupted natural migration among drainages (Magurran, 2005). Southern Trinidad is remote from both the drainages in the north (Figure 1), and it appears that there is little or no migration between the south and north areas based on the marked divergence of 720 single-nucleotide polymorphisms (Willing *et al.*, 2010).

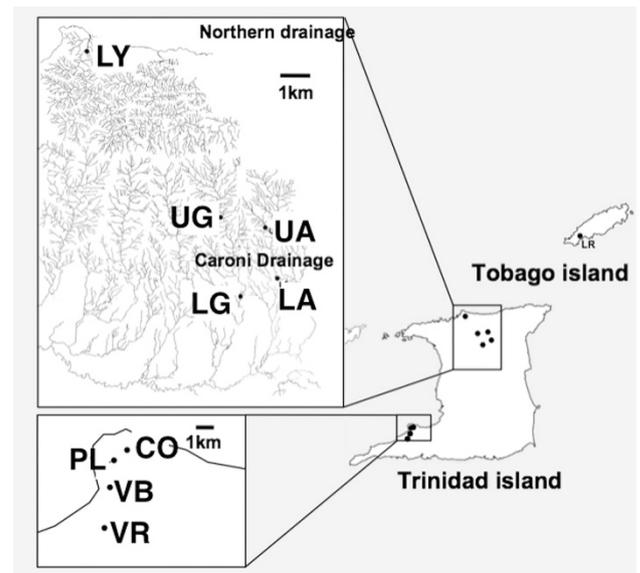


Figure 1 Locations of the sampling sites in Trinidad and Tobago. Upper Aripo (UA), Lower Aripo (LA), Upper Guanapo (UG) and Lower Guanapo (LG) are tributaries of the Caroni Drainage. One population, Lower Yarra (LY), came from a tributary of the Northern Drainage. Pitch Lake (PL), Visigney Beach (VB), Coffee (CO) and Vance River (VR) are located in southern Trinidad. Lambeau River (LR) is in Tobago.

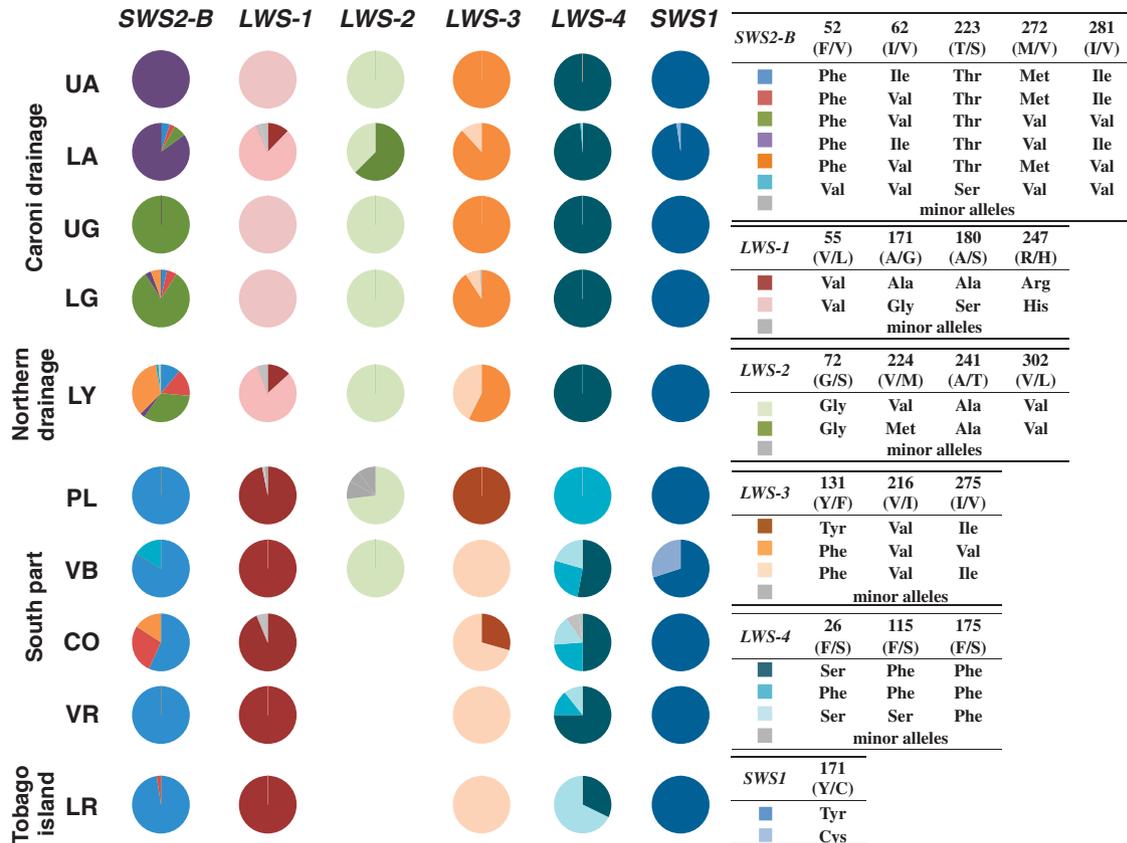


Figure 2 Allele frequencies of 6 opsin genes, that is, *SWS2-B*, *LWS-1*, *LWS-2*, *LWS-3*, *LWS-4* and *SWS1*, in the 10 guppy populations from Trinidad and Tobago. The different colors shown in the pie charts indicate different alleles that could be distinguished by non-Syn nucleotide differences. The tables on the right indicate the sites with amino acid variations. The number indicates the amino acid site. In *LWS-1*, amino acid 180 is one of the five sites known to affect the absorption spectra of opsins.

Aquatic light environments

Aquatic environmental data were collected at the sampling sites in Trinidad during the dry season in 2003. These measurements included the percentage canopy coverage that was determined by capturing images using a 35 mm wide angle lens with the camera pointed straight up while standing in the center of the sampling location. The percentage shade covering the water around noon was estimated by two independent observers. The dissolved oxygen (DO) and conductivity were measured using a YSI 55 multimeter probe (YSI Inc., Yellow Springs, OH, USA).

The canopy cover, shade and DO are appropriate indices for assessing aquatic light environments. Direct light spectrum data were not available in the present study. It should also be noted that light measurements such as the light spectrum and intensity per site vary greatly depending on the cloud cover and time of day, whereas the canopy cover does not vary temporally and has a significant effect on the light strength and spectrum in aquatic environments. The DO level is also closely related to the level of photosynthesis, the type of phytoplankton present and light penetration that affect the light intensity and color of aquatic environments (Twomey *et al.*, 2009). Thus, canopy cover and DO are appropriate and stable proxies for the light regime in aquatic environments.

Determination of the opsin nucleotide sequences

We isolated genomic DNA based on the acetyltrimethylammonium bromide protocol or using a Qiagen DNeasy96 kit (Qiagen, Crawley, UK). We designed PCR primers (Supplementary Table S1) based on the untranslated regions of six opsin genes (*SWS2-B*, *LWS-1*, *LWS-2*, *LWS-3*, *LWS-4* and *SWS1*), according to published sequences of Cumana guppy opsin genes (Watson *et al.*, 2011) (GenBank accession numbers HM540108 and HM540107) and our unpublished data. The complete coding sequences of these opsin genes were amplified using these primers. The amplification reactions followed a PCR

protocol with a volume of 50 µl that contained 2.5 µl of the genomic DNA template, 5.0 µl 10 × ExTaq Buffer, 1 µM dNTP mixture, 0.5 µM MgCl₂, 1 unit ExTaq (Takara Biotechnology Co., Tokyo, Japan) and the forward and reverse primers. The amplification conditions were as follows: 3 min at 95 °C; 35 cycles of 30 s at 94 °C; 30 s at 60 °C; and 4 min (*LWS-1* and *LWS-3*), 6 min (*LWS-2*), 3 min (*LWS-4*) or 2 min (*SWS1* and *SWS2-B*) at 72 °C, followed by 7 min at 72 °C. All the PCR products were purified by polyethylene glycol precipitation. In our analyses, we excluded *LWS-2* from the CO, VR and LR populations because the PCR amplifications of these samples were not successful.

The sequencing reactions were performed using a BigDye Terminator v3.1 Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Tokyo, Japan) with the PCR primers and additional internal primers. The sequences were determined in both strands using ABI3130 and ABI3170 Genetic Analyzer systems (Applied Biosystems). The sequences obtained were edited and aligned using CLUSTALX (Larkin *et al.*, 2007), MAFFT (Katoh and Toh, 2010) and Se-AL (Rambaut, 1996). The haplotype frequencies were estimated using PHASE 2.1 that uses Bayesian algorithms (Stephens *et al.*, 2001; Stephens and Donnelly, 2003) in DnaSP ver. 5 (Librado and Rozas, 2009). Deviation from a Hardy–Weinberg equilibrium was tested using Arlequin ver. 3.1 (Excoffier *et al.*, 2005). The alignments of all types of coding sequences are shown for each gene in Supplementary Figure S1. GenBank accession numbers were assigned to the most frequent sequence type for each gene (for *LWS-1*, one each of A-type and S-type are given that were distinguished at residue 180): A-type *LWS-1* (AB748984), S-type *LWS-1* (AB748985), *LWS-2* (AB748986), *LWS-3* (AB748987), *LWS-4* (AB748988), *SWS2-B* (AB748990) and *SWS1* (AB748989).

Determination of reference sequences

We used non-opsin reference genes (*ACTB_ORYLA*, *ADK*, *KCND2*, *MAB21L1*, *NABI*, *RBM4* and *SIX3*) to detect the natural selection acting on opsin genes

(Tezuka *et al.*, 2012). We selected 8 or 9 individuals randomly from each of the 10 populations and analyzed their reference sequences. We used the method developed by Tezuka *et al.* (2012) to obtain the reference sequences. Based on genomic information from the fully sequenced organisms, we designed primer pairs that we expected to amplify DNA fragments (300–700 bp) of the seven genes. Among these fragments, we selected primer pairs for DNA amplification of *P. reticulata* sequences where the sequence identities between the corresponding regions of the fully sequenced organisms were 75–90% that corresponded to the average identities of the exonic and/or intronic parts of all the genes in the genome (Tezuka *et al.*, 2012). Thus, the selected DNA fragments could be random samples from the whole genomes with respect to the degree of sequence conservation; therefore, we could only use these fragments as references for comparison with the focal opsin genes.

The haplotype frequencies were estimated using PHASE 2.1 (Stephens *et al.*, 2001; Stephens and Donnelly, 2003) in DnaSP ver. 5 (Librado and Rozas, 2009).

Summary nucleotide variation statistics

We calculated summary statistics for the reference genes and opsin genes for each population (Supplementary Tables S2 and S3) using DnaSP version 5 (Librado and Rozas, 2009). The ratio of nonsynonymous (non-Syn) to synonymous (Syn) substitutions in the six opsin genes was examined using Fisher's exact test (<http://www.physics.csbsju.edu/stats/exact.html>). The nucleotide diversity (π) is the average number of pairwise nucleotide differences per site that can be a direct estimator of the population mutation rate, θ , in the infinite site model under the standard Wright–Fisher model (Tajima, 1983). The population mutation rate is defined as $\theta = 4N\mu$, where N is the effective population size and μ is the mutation rate per nucleotide site per generation. The number of polymorphic (segregating) sites among sequences (S) also provides another estimator of θ when adjusted using the sample size (n) and the length of sequenced region (L). This estimate is referred to as θ_W (Watterson's θ ; Watterson, 1975). Tajima's D evaluates the difference between π and θ_W that are expected to be identical under the standard Wright–Fisher model (Tajima, 1989).

Deviations from Hardy–Weinberg heterozygosity within populations were expressed as $F_{IS} = 1 - Ho/He$, where Ho is the observed heterozygosity and He is the expected heterozygosity (Supplementary Tables S4 and S5). We calculated Ho and He using Arlequin ver. 3.1 (Excoffier *et al.*, 2005). Differences in F_{IS} among populations and genes were tested using a General Linear Model, where F_{IS} for each gene and population was the response variable, and population and gene were random factors. Random effects models are used when the treatments are not fixed, that is, when the various factor levels are sampled from a larger population. Genetic differentiation, F_{ST} , was calculated as $(\pi_{Total} - \pi_{Within})/\pi_{Total}$ where π_{Total} and π_{Within} are the π values for all populations and each population, respectively.

Coalescent simulations of Tajima's D test for balancing selection

We used Tajima's D statistic as a test of neutrality to detect balancing selection in the opsin genes. Given that Tajima's D is sensitive to demographic effects, we first estimated the parameters of a rough demographic model based on the non-opsin reference regions and we then obtained the null distribution of Tajima's D statistic. The procedure is similar to that reported by Hiwatashi *et al.* (2010) and Innan (2006).

We performed a coalescent simulation to infer demography. For the coalescent simulation, we used the 'ms' program (Hudson, 2002) that can simulate nucleotide polymorphism patterns under the infinite site model with population structure. We tested the hypothesis that balancing selection may operate on opsin genes that would probably make Tajima's D value positive. Therefore, to detect balancing selection, a major concern is population structure that can also result in a positive Tajima's D value. There appeared to be some fragmentation within the sample populations because most of the Tajima's D values were positive in the reference regions (Supplementary Table S2). Our assumption of constant population size (in the upland sites) appeared to be reasonable because Barson *et al.* (2009) found that even small upland guppy populations were stable and approached mutation–drift equilibrium. We performed coalescent simulations for each local population, where we assumed

that there were two subpopulations and we allowed migration between them. The migration rate was set so it was consistent with the observed polymorphism data characterized by π and θ_W (see below for details). This treatment should make the test for balancing selection conservative.

The simulation required two parameters: the population mutation and migration rates ($\theta = 4N\mu$ and $M = 4Nm$). We used a coalescent-based rejection-sampling algorithm to estimate these two parameters from the polymorphism data in the reference regions (summarized by π and θ_W). This approach was applied to each population where several reference regions were available (six for the PL and LY populations, and seven for the others; Supplementary Tables S2 and S5). The length of the sequence was set to the average of the six/seven reference regions examined, that is, 489 bp for PL and LY and 482 bp for other populations (Supplementary Table S2). The number of sequences was set to twice the total number of individuals examined, that is, 16 sequences because the reference regions were all autosomal (Supplementary Table S2). Some populations contained the reference region(s) in 14 samples (PL, VB, CO and VR), 18 samples (LA, UA, LG, UG and LR) or both (LY), but we used 16 as the overall average. These data were used to compute π and θ_W (Supplementary Table S2).

To estimate the values of θ and M for the observed values of π and θ_W (Supplementary Table S6), we used a simulation-based rejection-sampling algorithm, where random polymorphism patterns were generated using the 'ms' program. As mentioned above, we simulated two equal-sized subpopulations and estimated θ and M . In each replicate, θ and M were assigned randomly from the intervals, 0 to 2.00×10^{-3} and 0–10, respectively. Using these parameters, coalescent simulations were performed for six (or seven) independent regions to calculate π and θ_W . π_{ave}' and θ_{Wave}' were the average values of π and θ_W from the simulated data. Replicates were accepted if the values of π_{ave}' and θ_{Wave}' were within $\pm 5\%$ of the mean of the observed π and θ_W values (Supplementary Table S6). The distribution of a significant number of accepted pairs of θ and M was considered to be approximately proportional to the likelihood. This process was repeated until 10 000 accepted values of θ' and M' were accumulated.

Next, we tested whether the observed π and θ_W for the opsin genes could be explained using a neutral model by assuming these estimated parameters (Supplementary Table S6). The simulation was performed based on the condition of the density distributions of θ and M that were estimated from the six/seven reference regions, and we obtained the null distributions of the summary statistics for the opsin genes. The length of the sequence was set to the total length of the opsin genes examined (1071 bp for *LWS-1* and *LWS-3*, 1074 bp for *LWS-2* and *LWS-4*, 1441 bp for *SWS1* and 1821 bp for *SWS2*; Supplementary Table S3). The number of sequences was set to the total number of chromosomes examined (Supplementary Table S3). For each of the 10 000 θ and M values, we determined the polymorphism distribution and used it to calculate Tajima's D (Supplementary Figure S2). We then used the distribution of these Tajima's D values as a null distribution to test the neutrality of the opsin genes. We obtained a P -value for the proportion of simulation runs where the value of Tajima's D was greater than the observed value (Supplementary Figure S2).

Differences in Tajima's D among populations and genes were tested using a General Linear Model, where Tajima's D for each gene and population was the response variable (60 values), and population and gene were random factors.

Coalescent simulations for F_{ST} analysis of divergent selection

The idea of detecting local adaptation using F_{ST} was first proposed by Lewontin and Krakauer (1973). They proposed that the null distributions of F_{ST} in the reference loci can be approximated as χ^2 distributions. Therefore, if the mean F_{ST} value is estimated from multiple loci, it is possible to test whether a specific locus is an outlier of the null distribution because of local adaptation (the so-called 'Lewontin–Krakauer test') (Beaumont, 2005). However, given the increases in the amounts of data and computational power, recent studies tend to use empirical distributions of putatively neutral loci (Akey *et al.*, 2002) or coalescent simulations for null distributions (Beaumont and Nichols, 1996).

In the present study, we performed coalescent simulations to evaluate the local adaptation of opsin genes. Again, we used a simulation-based approach,

where we estimated the demographic parameters from the non-opsin regions initially and the null distribution for the opsin genes was obtained based on the condition of the estimated parameters (Supplementary Tables S4 and S5). We calculated the mean F_{ST} values of six reference loci with a randomly varying population migration rate ($4Nm$) from an arbitrary range of 0–3, using the 'ms' program (Hudson, 2002). We assumed an island model with 10 subpopulations with equal migration and 16 samples per subpopulation. We set θ as 1 because this does not affect the F_{ST} values. We accepted simulation runs if the mean F_{ST} value was within $\pm 1\%$ of the observed mean F_{ST} value. We repeated the simulation until we obtained 10 000 samples. The posterior distribution of $4Nm$ that we obtained is shown in Supplementary Figure S3A. Next, we used coalescent simulations based on the posterior distribution of $4Nm$ to confirm that the F_{ST} values of the reference loci were within the null distribution (Table 2 and Supplementary Figure S3B). The sample sizes were different; therefore, we calculated the null distribution of each opsin gene and obtained P -values that was similar to the test for balancing selection (Supplementary Figures S3C–G). *LWS-2* was excluded from this analysis because these sequences could not be determined in the CO, VR and LR populations.

Linkage disequilibrium (LD) among *SWS2-B*, *LWS-1*, *LWS-2* and *LWS-3* in six populations

We confirmed the existence of a LD between the *SWS2-B*, *LWS-1*, *LWS-2* and *LWS-3* opsin genes that are located in close proximity on the same chromosome. The LD patterns (LOD and D') were calculated using Haploview 4.2 (Barrett *et al.*, 2005; Barrett, 2009).

Relationships between opsin divergence and environmental differences between the sites

We used Mantel tests to examine the relationships between F_{ST} and the environmental differences between the sites. We used four aquatic environment variables: the percentage of canopy over the water surface, shade, DO and conductivity. We examined nine sampling sites in Trinidad because no data were available for Tobago (Supplementary Table S7). Next, distance matrices were constructed between the sites by calculating the pairwise differences in the values of the environmental variables between sites. Mantel tests were conducted between the pairwise F_{ST} values and the differences in each environmental variable between the sites. The geographic distance between the sites was also examined to account for the effects of gene flow. The P -values were adjusted for the inflated type I error rate because of multiple comparisons using the false discovery rate procedure. In addition, we incorporated geographic distances in the partial Mantel tests for environmental variables that were significantly related to F_{ST} , thereby controlling for the effect of genetic dissimilarity caused by the genetic isolation of populations from distinct drainages or regions (the Caroni and Northern Drainages, and Southern Trinidad).

RESULTS

Allelic differentiation of M/LWS opsins for spectral sensitivity

The amino acid compositions of the spectrally effective sites, that is, 180, 197, 277, 285 and 308 of the four M/LWS-type opsin genes, were as follows: Ala, His, Tyr, Thr and Ala (designated Ala/His/Tyr/Thr/Ser), respectively, in *LWS-1*; Pro/His/Phe/Ala/Ser in *LWS-2*; and Ser/His/Tyr/Thr/Ser in *LWS-3* and *LWS-4*, as reported previously (Watson *et al.*, 2011). The composition was invariant within and between populations except residue 180 of *LWS-1* where an Ala/Ser polymorphism was found. Ala was the major allele in southern Trinidad and Tobago, whereas Ser was the major allele in the Northern and Caroni drainage areas of Trinidad (Figure 2 and Supplementary Table S8).

Overall nucleotide variation in opsin genes

The gene-level diversity measures for the opsin genes and non-opsin reference genes are shown in Supplementary Tables S4 and S5, respectively. The genotype frequencies of opsin genes in some

populations deviated significantly from the Hardy–Weinberg equilibrium (mean (\pm s.e.) opsin F_{IS} = 0.189 (\pm 0.035), Supplementary Table S4), whereas the reference genes were at Hardy–Weinberg equilibrium (mean (\pm s.e.) F_{IS} = 0.050 (\pm 0.053), Supplementary Table S5). The heterozygote deficiency was significantly higher for the opsin genes compared with the non-opsin reference genes (Mann–Whitney U -test, W = 1851.5, P = 0.020; Supplementary Figure S4).

There were 4 non-Syn and 21 Syn polymorphic sites in *LWS-1*, 4 non-Syn and 4 Syn polymorphic sites in *LWS-2*, 3 non-Syn and 14 Syn polymorphic sites in *LWS-3* and 3 non-Syn and 9 Syn polymorphic sites in *LWS-4*. The short wavelength opsins were also polymorphic, where the *SWS1* coding region had 1 non-Syn and 4 Syn polymorphic sites, whereas the *SWS2-B* coding region had 5 non-Syn and 29 Syn polymorphic sites. There was no significant difference in the ratio of non-Syn to Syn substitutions among the six opsin genes (Fisher's exact test, P = 0.243).

Figure 2 shows the allele frequencies of the six opsin genes in the 10 populations in terms of the amino acid haplotype. The allele frequencies differed among populations and the differences were notably larger between the Northern/Caroni drainages of Trinidad (LY, UA, LA, UG and LG) and the other populations (Figure 2). The opsin genes in the downstream populations (LA, LG and LY) were more polymorphic than those in the upstream population (UA and UG; Figure 2 and Supplementary Tables S4 and S5) that is consistent with the higher level of genetic drift and genetic isolation found in upland regions (Barson *et al.*, 2009; Willing *et al.*, 2010). The *SWS2-B* opsin gene was highly polymorphic and the allele compositions differed among the streams.

Test for balancing selection in opsin genes

The coalescent simulation showed that the observed Tajima's D values of *SWS2-B* in CO and *LWS-3* in LY were significantly larger than those of the reference regions (P = 0.001 in CO and 0.012 in LY; Table 1 and Supplementary Figure S5). Those of the reference genes were within the distribution (Supplementary Figure S5). However, after correcting for multiple tests using Bonferroni's method, the deviations in these opsin genes were not statistically significant (P = 0.072 in CO and 0.649 in LY). Tajima's D did not differ significantly between the opsin genes (Friedman rank sum test, Friedman χ^2 = 3.03, d.f. = 5, P = 0.695), but there were significant differences in the value of Tajima's D between populations for the opsin genes (Friedman χ^2 = 19.60, d.f. = 9, P = 0.021; Supplementary Figure S6).

Testing for divergent selection in opsin genes

The mean F_{ST} for all populations and the pairwise F_{ST} values of the opsin genes and the non-opsin reference genes are shown in Table 2 and Supplementary Table S9, respectively. We found that the guppy populations appeared to be more diverged with respect to the opsin genes (*SWS2-B*, *LWS-1*, *LWS-3* and *LWS-4*) than the non-opsin reference genes (Supplementary Figure S3) (*SWS2-B*: mean F_{ST} = 0.596, P = 0.037; *LWS-1*: mean F_{ST} = 0.810, P < 0.001; *LWS-3*: mean F_{ST} = 0.735, P = 0.002; *LWS-4*: mean F_{ST} = 0.637, P = 0.018; *SWS1*: mean F_{ST} = 0.549, P = 0.085). The P -values of *LWS-1* and *LWS-3* remained significant after correction for multiple testing (*LWS-1*: P < 0.001, *LWS-3*: P < 0.012). More importantly, we found that the observed mean value of the opsin genes (mean F_{ST} = 0.693) deviated significantly from that of the reference loci (mean F_{ST} = 0.419, P < 0.001, Supplementary Figure S3H). Note that to be conservative, we assumed three loci in the coalescent simulation

Table 1 Sample sizes (*n*), Tajima's *D* and *P*-values for the six opsin genes

Population	SWS2-B	LWS-1	LWS-2	LWS-3	LWS-4	SWS1
UA						
<i>n</i>	70	66	66	58	66	44
Tajima's <i>D</i>	0	0.772	0	0	-1.074	-1.653
<i>P</i> -values	0.528	0.279	0.495	0.484	0.881	0.988
LA						
<i>n</i>	102	90	82	82	86	94
Tajima's <i>D</i>	0.137	0.624	1.11	0.663	-0.289	-0.463
<i>P</i> -values	0.434	0.269	0.158	0.259	0.595	0.668
UG						
<i>n</i>	96	94	70	72	84	76
Tajima's <i>D</i>	0	-1.035	0	0	0	0
<i>P</i> -values	0.469	0.867	0.457	0.456	0.451	0.465
LG						
<i>n</i>	88	76	72	98	82	74
Tajima's <i>D</i>	-0.802	-1.321	-1.246	-0.579	-1.599	1.566
<i>P</i> -values	0.78	0.943	0.927	0.697	0.989	0.079
LY						
<i>n</i>	72	56	42	62	76	78
Tajima's <i>D</i>	-1.43	0.684	-0.001	2.857	0.898	-1.055
<i>P</i> -values	0.968	0.28	0.586	0.012	0.231	0.876
PL						
<i>n</i>	184	96	90	72	156	94
Tajima's <i>D</i>	-1.32	-1.95	-0.366	0	-1.053	-0.474
<i>P</i> -values	0.954	0.999	0.671	0.371	0.907	0.678
VB						
<i>n</i>	36	32	10	20	32	20
Tajima's <i>D</i>	0.646	1.039	-0.762	1.036	1.14	1.404
<i>P</i> -values	0.27	0.175	0.767	0.173	0.151	0.102
CO						
<i>n</i>	44	46	NA	32	44	36
Tajima's <i>D</i>	3.087	-0.282		1.062	1.66	1.625
<i>P</i> -values	0.001	0.596		0.154	0.06	0.062
VR						
<i>n</i>	20	24	NA	16	28	12
Tajima's <i>D</i>	-1.513	0.525		0	-0.822	0.472
<i>P</i> -values	0.955	0.289		0.468	0.788	0.322
LR						
<i>n</i>	82	80	NA	74	58	80
Tajima's <i>D</i>	-1.698	-0.382		-1.013	0.484	-0.766
<i>P</i> -values	0.989	0.619		0.846	0.296	0.768

Abbreviations: CO, Coffee; LA, Lower Aripo; LG, Lower Guanapo; LR, Lambeau River; LY, Lower Yarra; NA, the sequence data could not be obtained; PL, Pitch Lake; UA, Upper Aripo; UG, Upper Guanapo; VB, Visigney Beach; VR, Vance River.
The *P*-values were not adjusted for multiple comparisons.

instead of five loci because of the strong LD between three opsin genes (Supplementary Figure S7). The results suggest that the opsin genes have been subject to divergent selection.

Relationships between the environmental variables and opsin genes

The pairwise F_{ST} values and the differences in the DO values between the sampling sites were significantly correlated for *LWS-1* (Mantel test: $P=0.019$) and *LWS-3* (Mantel test: $P<0.001$; Figure 3 and Supplementary Table S10). The pairwise F_{ST} values of *LWS-1* and *LWS-3* were marginally significant for the geographic distances (Mantel tests: *LWS-1*, $P=0.054$ and *LWS-3*, $P=0.049$). The partial Mantel test showed that the positive relationships between F_{ST} and DO remained significant after controlling for the geographic distances (*LWS-1*, $r=0.330$, $P=0.024$ and *LWS-3*, $r=0.616$, $P<0.001$).

Table 2 Average F_{ST} values for the opsin and non-opsin reference genes

Gene	F_{STave}	All F_{STave}
Opsin		
<i>SWS2-B</i>	0.596	
<i>LWS-1</i>	0.810	
<i>LWS-2</i>	0.651	
<i>LWS-3</i>	0.735	
<i>LWS-4</i>	0.637	
<i>SWS1</i>	0.549	0.693
Non-opsin		
<i>ACTB_ORYLA</i>	0.226	
<i>ADK</i>	0.401	
<i>KCND2</i>	0.386	
<i>MAB21L1</i>	0.465	
<i>NABI</i>	0.516	
<i>RBM4</i>	0.059	
<i>SIX3</i>	0.444	0.419

The F_{ST} average (F_{STave}) values of *LWS-2* and *RBM4* were calculated using seven and eight populations, respectively, because data were not available for some populations. All the average F_{ST} values were calculated using the F_{ST} values of the opsin genes and non-opsin genes, excluding *LWS-2* and *RBM4*.

DISCUSSION

In the present study, we examined the nucleotide variation in the four M/LWS opsin genes as well as the blue-sensitive *SWS2-B* and UV-sensitive *SWS1* opsin genes in 10 guppy populations in various light environments in Trinidad and Tobago. For the first time, we discovered a potential spectral variation (180 Ser/Ala) in *LWS-1* that differed at an amino acid site that is known to affect the absorption spectra of opsins. A coalescent simulation showed that the inter-population genetic differentiation of two opsin genes (*LWS-1* and *LWS-3*) was significantly higher than the neutral expectation. This genetic differentiation was related significantly to differences in the DO level. These results suggest that the diversity of the opsin genes among populations is driven significantly by natural selection and that guppies can adapt to various light environments by changes in their color vision.

Allelic differentiation of M/LWS opsins with respect to spectral sensitivity

In the present study, we detected an allelic polymorphism in the five-site composition at *LWS-1* (180/197/277/285/308 = Ala/His/Tyr/Thr/Ala, A-type, or Ser/His/Tyr/Thr/Ala, S-type) that differed at amino acid 180. Single mutations, that is, S180A, H197Y, Y277F, T285A and A308S, and double mutations, S180A/H197Y, in the vertebrate M/LWS opsins shift λ_{max} by -7, -28, -8, -15, -27 and 11 nm, respectively, that is known as the 'five-sites' rule (Yokoyama and Radlwimmer, 2001). In addition, the S180P mutation is reported to shift the λ_{max} of the lamprey LWS photopigment by almost 19 nm toward a shorter wavelength (Davies *et al.*, 2009). According to the five-site compositions of other species (Yokoyama and Radlwimmer, 2001; Davies *et al.*, 2009), the λ_{max} values of the S-type and the A-type *LWS-1* are expected to be 560 and 553 nm, respectively. The λ_{max} of *LWS-2* (Pro/His/Phe/Ala/Ser) is expected to be 518 nm. The λ_{max} of both *LWS-3* and *LWS-4* (Ser/His/Tyr/Thr/Ala, which is the same as the S-type *LWS-1*) is expected to be 560 nm.

We can speculate about an ancestral allele by comparing closely related species. In a previous study of the Cumaná guppy in Venezuela

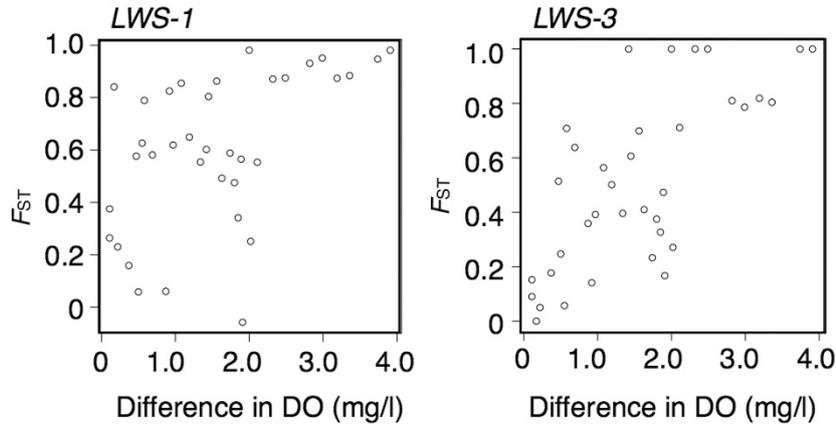


Figure 3 Relationships between the F_{ST} value and DO in *LWS-1* and *LWS-3*. The F_{ST} values are the pairwise values between the nine Trinidad populations. The DO values are the differences in the DO concentrations between the nine Trinidad populations. A Mantel test showed that both relationships were significant (*LWS-1*, $P=0.019$, *LWS-3*, $P<0.001$, Supplementary Table S10).

(Watson *et al.*, 2011), *LWS-1* was reported to be the A-type. In our study, the A-type was the major allele in southern Trinidad and in Tobago, whereas the S-type was the major allele in northern Trinidad (that is, the Northern Drainage and the Caroni Drainage). In the green swordtail (*Xiphophorus helleri*), a closely related species to the guppy, *LWS-1* was reported to be the S-type (Watson *et al.*, 2010). In a study of the four-eyed fish (*Anableps anableps*) (Owens *et al.*, 2009), a close outgroup of the guppy and the swordtail, multiple M/LWS-type opsin genes were also reported and residue 180 was Ser in all these genes. Taken together, these results may suggest that the A-type *LWS-1* allele was derived from the common ancestor of the Venezuelan and the Trinidad and Tobago guppies after their divergence from the swordtail. Thus, the S-type *LWS-1* allele found in the northern part of Trinidad could be an ancestral remnant. However, it is also possible that it was derived secondarily from *LWS-3* by gene conversion. Gene conversion has also been implicated in the generation of M/LWS gene variation in the guppy, the swordtail and the four-eyed fish (Owens *et al.*, 2009; Watson *et al.*, 2010, 2011).

The present study included samples from two populations (UA and LA) in the Aripo river that Archer and Lythgoe (1990) had also sampled and they detected sensory polymorphisms. In the Lower Aripo (LA) population, all four of the *LWS* opsin genes had amino acid polymorphisms (Figure 2). Archer and Lythgoe (1990) detected three major types of cone cells with different λ_{max} values (533, 548 and 572 nm) using MSP. The predicted λ_{max} values of *LWS-2* (518 nm; Pro/His/Phe/Ala/Ser), A-type *LWS-1* (553 nm; Ala/His/Tyr/Thr/Ala) and *LWS-3/LWS-4/S-type LWS-1* (560 nm; Ser/His/Tyr/Thr/Ala) may correspond to the three major types detected by MSP. However, Archer and Lythgoe (1990) also reported that different individuals had various combinations of these cell types where individuals had one, two or three types. In addition, a group of cells had a broad range of λ_{max} values at ~ 548 nm. These findings cannot be explained by allelic combinations of the three predicted λ_{max} values of the M/LWS opsins determined in the present study. Thus, there may be further variation in λ_{max} among the four M/LWS loci, and there is a further possibility of allelic differentiation in the four loci related to λ_{max} because of amino acid variations other than the five sites as well as different expression patterns. The sensory variation could also be related to interindividual variation in the relative expression levels among opsin loci, as found in the Venezuelan guppies (Rennison *et al.*, 2011). Direct measurements of λ_{max} for the

reconstituted photopigments and quantification of the expression of the alleles are necessary to resolve these questions.

We also found genetic polymorphisms at non-Syn sites in *SWS2-B* and *SWS1*. This suggests that there could be phenotypic variation in short wavelength color vision, although phenotypic variation has not been reported previously for short wave-sensitive cells using MSP. In total, we detected 6 non-Syn variable sites out of a total of 39 polymorphic sites in the two short wave opsin-coding regions, compared with 14 out of 67 sites in the combined M/LWS-coding regions. The ratio of Syn to non-Syn substitutions did not differ significantly between the long and short wavelength opsin genes in the samples analyzed. Many organisms use ultraviolet radiation for individual discrimination, communication and direction recognition (Losey *et al.*, 1999; Honkavaara *et al.*, 2002; Siitari *et al.*, 2002). Female guppies use ultraviolet radiation from the male body during mate choice (Kodric-Brown and Johnson, 2002; Smith *et al.*, 2002). At present, we are confirming the differences in λ_{max} based on sequence variation in *SWS*.

Balancing selection for opsin genes

The coalescent simulations showed that the observed Tajima's D values of *LWS-3* at LY and *SWS2-B* at CO were apparent outliers in the distribution expected from the reference genes, but they were not significant after correcting for multiple comparisons. The results showed that there was no evidence for balancing selection at any loci. Thus, our results suggest that there is no evidence for overdominance or negative frequency-dependent selection within subpopulations with respect to the opsin genes. The observed heterozygote deficiency observed at opsin genes in some lowland populations suggests the opposite (see below).

Divergent selection for opsin genes

The overall genetic divergence among the populations (F_{ST}) for the opsin genes was significantly greater than that for the non-opsin reference genes (Table 2 and Supplementary Figure S3). In particular, the coalescent simulation indicated that the F_{ST} values for the *LWS-1* and *LWS-3* loci were significantly higher than those for the non-opsin reference genes. One of these *LWS* loci may be significantly divergent among populations because there was a very high level of LD among the loci (Supplementary Figure S7).

This observation is consistent with divergent selection operating on opsin genes, particularly *LWS* loci, resulting in local adaptation. In

addition, we observed significant heterozygote deficiency for opsin genes that was particularly prominent in the downstream populations (the LA and CO populations were significant with an α of 5%; Supplementary Table S1). In contrast, the reference genes were at Hardy–Weinberg equilibrium (Supplementary Table S4). The most consistent explanation for this observation is that divergent selection acts on the opsin genes that could have changed the allele frequencies in the tributaries that flow into the downstream populations. These migrants may be homozygous for alleles that are rare in the recipient downstream population, resulting in the observed heterozygote deficiency. Consistent with this explanation is the study by Barson *et al.* (2009), who showed that the lowland Caroni populations also had a heterozygote deficit at highly polymorphic microsatellite markers that they explained by the inflow of migrants from upstream populations when the rivers were in spate during the wet season rains (van Oosterhout *et al.*, 2007; McMullan and van Oosterhout, 2012).

The allele frequencies of *SWS2-B*, *LWS-1*, *LWS-2* and *LWS-3* differed among populations, and the differences between northern Trinidad (Caroni and Northern drainages) and the other populations were notably larger (Figure 2). In the northern populations, there were large variations in the allele frequencies, particularly for *SWS2-B*. Three to five non-Syn sites caused the differences among the alleles. The relatively minor alleles in northern Trinidad were dominant in southern Trinidad for *SWS2-B*, *LWS-1* and *LWS-3*. Given the reduced amount of variation in the reference genes, this suggests that divergent selection caused genetic differentiation among the populations and that different alleles of the opsin genes may adapt to different environmental conditions.

Our results showed that the divergence of opsin genes was related significantly to differences in one environmental variable, that is, DO, particularly for two opsin genes (*LWS-1* and *LWS-3*) with loci that were subject to statistically significant divergent selection. In addition, the differences in the DO levels between populations were still significantly related to the F_{ST} values of the opsin loci even when the effects of the geographic distances between populations were removed. Therefore, differences in DO may be related to the factors that cause divergent selection, independently of the effects of gene flow and divergent environmental factors related to geographic differences.

DO level may be related to eutrophication (the amount and types of phytoplankton) that affects both the water color and light penetration. In addition, changes in the DO level may cause changes in the suitable depth of aquatic environments (Vonlanthen *et al.*, 2012), thereby leading to water color variation in guppy habitats. We found that the frequency of S-type *LWS-1* (560 nm) increased significantly with the DO level (General Linear Model, binomial, logit link, $P = 1.49 \times 10^{-19}$). In *LWS-3*, a major allele (FVV, Figure 2) increased significantly with the DO level (General Linear Model, binomial, logit link, $P = 7.90 \times 10^{-9}$). These results suggest that long wave-sensitive genotypes may be favored by a higher DO level. There were variations in the *LWS* loci between the northern and southern populations in Trinidad. Unlike the guppies from the Northern and Caroni Drainages, those from Pitch Lake in southern Trinidad were exposed to direct sunlight, high water temperatures and the sediment produced a black background color (that is, pitch) (Kenny, 1995). However, the differences in the canopy cover and shade level between the sites were not highly related to opsin divergence; therefore, light intensity may not be a selective factor. Further studies are required to identify the selective agents that act on different opsin genes and genotypes.

In conclusion, we found considerable amounts of allelic variation of some opsin genes relative to reference genes in wild populations of

guppies, including one amino acid at a known spectral tuning site. The allelic variations at the two loci, *LWS-1* and *LWS-3*, were subject to divergent selection among the populations, and the DO level of the aquatic environment associated may be related to the divergent selective factors. Thus, it is necessary to explore spectral variation among the opsin alleles to increase our understanding of the functionality of allelic variants.

DATA ARCHIVING

The sequence data have been submitted to GenBank under accession numbers AB748984–AB748990.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank J Cable for providing the guppy fins, A Shoji, J Yokoyama and N Matsushima for determining the opsin gene sequences, T Makino for designing the primers for the reference genes and Y Terai for commenting on the manuscript. MK was supported by a Grant-in-Aid for Scientific Research (21370007) from the Japan Society for the Promotion of Science (JSPS) and partly by the Global COE Program ‘Centre for ecosystem management adapting to global change’ (J03) of the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT). SK was supported by a Grant-in-Aid for Scientific Research A (22247036), a Grant-in-Aid for Challenging Exploratory Research (23657164) from JSPS and Grants-in-Aid for Scientific Research on Priority Areas ‘Comparative Genomics’ (20017008) and ‘Cellular Sensor’ (21026007) from MEXT. MY was supported by a Research Fellowship from the Japan Society for the Promotion of Science (JSPS) for Young Scientists (21-7611). CvO and MM were funded by the Earth & Life Systems Alliance (ELSA). We thank M Hattori and K Oshima for support with nucleotide sequencing through a Grant-in-Aid for Scientific Research on Innovative Areas ‘Genome Science’ from Ministry of Education, Culture, Sports, Science and Technology of Japan (No. 221S0002).

-
- Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* **12**: 1805–1814.
- Archer SN, Endler JA, Lythgoe JN, Partridge JC (1987). Visual pigment polymorphism in the guppy *Poecilia reticulata*. *Vision Res* **27**: 1243–1252.
- Archer SN, Lythgoe JN (1990). The visual pigment basis for cone polymorphism in the guppy, *Poecilia reticulata*. *Vision Res* **30**: 225–233.
- Barrett JC (2009). Haploview: visualization and analysis of SNP genotype data. *Cold Spring Harb Protoc* **2009**: pdb.ip71.
- Barrett JC, Fry B, Maller J, Daly MJ (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**: 263–265.
- Barson NJ, Cable J, van Oosterhout C (2009). Population genetic analysis of microsatellite variation of guppies (*Poecilia reticulata*) in Trinidad and Tobago: evidence for a dynamic source-sink metapopulation structure, founder events and population bottlenecks. *J Evol Biol* **22**: 485–497.
- Beaumont MA (2005). Adaptation and speciation: what can F_{ST} tell us? *Trends Ecol Evol* **20**: 435–440.
- Beaumont MA, Nichols RA (1996). Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc* **263**: 1619–1626.
- Breden F (2006). Guppies. *Curr Biol* **16**: R865–R866.
- Brooks R (2002). Variation in female mate choice within guppy populations: population divergence, multiple ornaments and the maintenance of polymorphism. *Genetica* **116**: 343–358.
- Carvalho GR, Shaw PW, Magurran AE, Seghers BH (1991). Marked genetic divergence revealed by allozymes among populations of the guppy *Poecilia reticulata* (Poeciliidae), in Trinidad. *Biol J Linn Soc* **42**: 389–405.
- Davies WL, Collin SP, Hunt DM (2009). Adaptive gene loss reflects differences in the visual ecology of basal vertebrates. *Mol Biol Evol* **26**: 1803–1809.
- Endler JA (1992). Signals, signal conditions, and the direction of evolution. *Am Nat* **139**: S125–S153.
- Endler JA, Houde AE (1995). Geographic variation in female preferences for male traits in *Poecilia reticulata*. *Evolution* **49**: 456–468.
- Excoffier L, Laval G, Schneider S (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform* **1**: 47–50.

- Hiwatashi T, Okabe Y, Tsutsui T, Hiramatsu C, Melin AD, Oota H *et al.* (2010). An explicit signature of balancing selection for color-vision variation in new world monkeys. *Mol Biol Evol* **27**: 453–464.
- Hoffmann M, Tripathi N, Henz SR, Lindholm AK, Weigel D, Breden F *et al.* (2007). Opsin gene duplication and diversification in the guppy, a model for sexual selection. *Proc R Soc B* **274**: 33–42.
- Honkavaara J, Koivula M, Korpimäki E, Siitari H, Viitala J (2002). Ultraviolet vision and foraging in terrestrial vertebrates. *Oikos* **98**: 505–511.
- Houde AE, Endler JA (1990). Correlated evolution of female mating preferences and male color patterns in the guppy *Poecilia reticulata*. *Science* **248**: 1405–1408.
- Houde AE, Hankes MA (1997). Evolutionary mismatch of mating preferences and male colour patterns in guppies. *Anim Behav* **53**: 343–351.
- Hudson RR (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337–338.
- Innan H (2006). Modified Hudson-Kreitman-Aguadé test and two-dimensional evaluation of neutrality tests. *Genetics* **173**: 1725–1733.
- Jacobs GH, Rowe MP (2004). Evolution of vertebrate colour vision. *Clin Exp Optom* **87**: 206–216.
- Katoh K, Toh H (2010). Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* **26**: 1899–1900.
- Kenny JS (1995). *Views from the Bridge: A Memoir on the Freshwater Fishes of Trinidad*. The University of the West Indies Press: Port of Spain.
- Kodric-Brown A, Johnson SC (2002). Ultraviolet reflectance patterns of male guppies enhance their attractiveness to females. *Anim Behav* **63**: 391–396.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.
- Laver CR, Taylor JS (2011). RT-qPCR reveals opsin gene upregulation associated with age and sex in guppies (*Poecilia reticulata*)—a species with color-based sexual selection and 11 visual-opsin genes. *BMC Evol Biol* **11**: 81.
- Lewontin RC, Krakauer J (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**: 175–195.
- Librado P, Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.
- Losey GS, Cronin TW, Goldsmith TH, Hyde D, Marshall NJ, McFarland WN (1999). The UV visual world of fishes: a review. *J Fish Biol* **54**: 921–943.
- Maan ME, Hofker KD, van Alphen JJ, Seehausen O (2006). Sensory drive in cichlid speciation. *Am Nat* **167**: 947–954.
- Magurran AE (2005). *Evolutionary Ecology: The Trinidadian Guppy*. Oxford University Press: New York.
- McMullan M, van Oosterhout C (2012). Inference of selection based on temporal genetic differentiation in the study of highly polymorphic multigene families. *PLoS ONE* **7**: e42119.
- Owens GL, Windsor DJ, Mui J, Taylor JS (2009). A fish eye out of water: ten visual opsins in the four-eyed fish, *Anableps anableps*. *PLoS ONE* **4**: e5970.
- Rambaut A (1996). Se-AL: Sequence Alignment Editor. <http://tree.bio.ed.ac.uk/software/seal/>.
- Rennison DJ, Owens GL, Allison WT, Taylor JS (2011). Intra-retinal variation of opsin gene expression in the guppy (*Poecilia reticulata*). *J Exp Biol* **214**: Pt 19 3248–3254.
- Sandkam BA, Joy JB, Watson CT, Gonzalez-Bendixsen P, Gabor CR, Breden F (2013). Hybridization leads to sensory repertoire expansion in a gynogenetic fish, the Amazon molly (*Poecilia formosa*): a test of the hybrid-sensory expansion hypothesis. *Evolution* **67**: 120–130.
- Schwartz AK, Hendry AP (2007). A test for the parallel co-evolution of male colour and female preference in Trinidadian guppies (*Poecilia reticulata*). *Evol Ecol Res* **9**: 71–90.
- Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HD, Miyagi R *et al.* (2008). Speciation through sensory drive in cichlid fish. *Nature* **455**: 620–626.
- Siitari H, Honkavaara J, Huhta E, Viitala J (2002). Ultraviolet reflection and female mate choice in the pied flycatcher, *Ficedula hypoleuca*. *Anim Behav* **63**: 97–102.
- Smith EJ, Partridge JC, Parsons KN, White EM, Cuthill IC, Bennett ATD *et al.* (2002). Ultraviolet vision and mate choice in the guppy (*Poecilia reticulata*). *Behav Ecol* **13**: 11–19.
- Stephens M, Donnelly P (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* **73**: 1162–1169.
- Stephens M, Smith NJ, Donnelly P (2001). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**: 978–989.
- Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.
- Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Terai Y, Mayer WE, Klein J, Tichy H, Okada N (2002). The effect of selection on a long wavelength-sensitive (LWS) opsin gene of Lake Victoria cichlid fishes. *Proc Natl Acad Sci USA* **99**: 15501–15506.
- Terai Y, Seehausen O, Sasaki T, Takahashi K, Mizoiri S, Sugawara T *et al.* (2006). Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biol* **4**: e433.
- Tezuka A, Matsushima N, Nemoto Y, Akashi HD, Kawata M, Makino T (2012). Comprehensive primer design for analysis of population genetics in non-model organisms. *PLoS ONE* **7**: e32314.
- Twomey LJ, Piehler MF, Paerl HW (2009). Priority parameters for monitoring of freshwater and marine systems and their measurement. In: Inyang II, Daniels JL (eds) *Environmental Monitoring*. EOLSS: Ontario, pp 318–338.
- van Oosterhout C, Joyce DA, Cummings SM, Blais J, Barson NJ, Ramnarine IW *et al.* (2006). Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution* **60**: 2562–2574.
- van Oosterhout C, Mohammed RS, Hansen H, Archard GA, McMullan M, Weese DJ *et al.* (2007). Selection by parasites in spate conditions in wild Trinidadian guppies (*Poecilia reticulata*). *Int J Parasitol* **37**: 805–812.
- Vonlanthen P, Bittner D, Hudson AG, Young KA, Müller R, Lundsgaard-Hansen B *et al.* (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* **482**: 357–362.
- Ward MN, Churcher AM, Dick KJ, Laver CRJ, Owens GL, Polack MD *et al.* (2008). The molecular basis of color vision in colorful fish: four long wave-sensitive (LWS) opsins in guppies (*Poecilia reticulata*) are defined by amino acid substitutions at key functional sites. *BMC Evol Biol* **8**: 210.
- Watson CT, Gray SM, Hoffmann M, Lubieniecki KP, Joy JB, Sandkam BA *et al.* (2011). Gene duplication and divergence of long wavelength-sensitive opsin genes in the guppy, *Poecilia reticulata*. *J Mol Evol* **72**: 240–252.
- Watson CT, Lubieniecki KP, Loew E, Davidson WS, Breden F (2010). Genomic organization of duplicated short wave-sensitive and long wave-sensitive opsin genes in the green swordtail, *Xiphophorus helleri*. *BMC Evol Biol* **10**: 87.
- Watterson GA (1975). On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7**: 256–276.
- Weadick CJ, Chang BSW (2007). Long-wavelength sensitive visual pigments of the guppy (*Poecilia reticulata*): six opsins expressed in a single individual. *BMC Evol Biol* **7**: S11.
- Willing EM, Bentzen P, van Oosterhout C, Hoffmann M, Cable J, Breden F *et al.* (2010). Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. *Mol Ecol* **19**: 968–984.
- Yokoyama S (2000). Molecular evolution of vertebrate visual pigments. *Prog Retin Eye Res* **19**: 385–419.
- Yokoyama S, Radlwimmer FB (1998). The “five-sites” rule and the evolution of red and green color vision in mammals. *Mol Biol Evol* **15**: 560–567.
- Yokoyama S, Radlwimmer FB (2001). The molecular genetics and evolution of red and green color vision in vertebrates. *Genetics* **158**: 1697–1710.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)