Functional Diversification of Lepidopteran Opsins Following Gene Duplication

Adriana D. Briscoe[1]

Department of Molecular and Cellular Biology, University of Arizona

A comparative approach was taken for identifying amino acid substitutions that may be under positive Darwinian selection and are correlated with spectral shifts among orthologous and paralogous lepidopteran long wavelength–sensitive (LW) opsins. Four novel LW opsin fragments were isolated, cloned, and sequenced from eye-specific cDNAs from two butterflies, *Vanessa cardui* (Nymphalidae) and *Precis coenia* (Nymphalidae), and two moths, *Spodoptera exigua* (Noctuidae) and *Galleria mellonella* (Pyralidae). These opsins were sampled because they encode visual pigments having a naturally occurring range of λ_{max} values (510–530 nm), which in combination with previously characterized lepidopteran opsins, provide a complete range of known spectral sensitivities (510–575 nm) among lepidopteran LW opsins. Two recent opsin gene duplication events were found within the papilionid but not within the nymphalid butterfly families through neighbor-joining, maximum parsimony, and maximum likelihood phylogenetic analyses of 13 lepidopteran opsin sequences. An elevated rate of evolution was detected in the red-shifted *Papilio* Rh3 branch following gene duplication, because of an increase in the amino acid substitution rate in the transmembrane domain of the protein, a region that forms the chromophore-binding pocket of the visual pigment. A maximum likelihood approach was used to estimate ω, the ratio of nonsynonymous to synonymous substitutions per site. Branch-specific tests of selection (free-ratio) identified one branch with $\omega = 2.1044$, but the small number of substitutions involved was not significantly different from the expected number of changes under the neutral expectation of $\omega = 1$. Ancestral sequences were reconstructed with a high degree of certainty from these data. Reconstructed ancestral sequences revealed several instances of convergence to the same amino acid between butterfly and vertebrate cone pigments, and between independent branches of the butterfly opsin tree that are correlated with spectral shifts.

Introduction

Opsins are members of the G protein–coupled receptor family of proteins characterized by seven membrane-spanning alpha helical domains. The transmembrane domains of the opsin form a binding pocket around a light-sensitive vitamin A–derived chromophore. Together, the opsin and the chromophore form a visual pigment expressed in the photoreceptor cells of an organism. Photoisomerization of the chromophore generates a cascade of signaling events in the photoreceptor cell that result in the generation of an electrical signal to the nervous system that light has been detected. Visual pigments have characteristic absorption maxima (λ_{max}) that are modulated by a handful of specific amino acid side groups that face the binding pocket of the chromophore (e.g., Merbs and Nathans 1993; Asenjo, Rim, and Oprian 1994). Isolated from the opsin, the chromophore has a characteristic λ_{max} in the ultraviolet. The process of gene duplication and sequence divergence of the opsin family has resulted in opsins that, in complex with the chromophore, form visual pigments that absorb throughout the visible light spectrum.

The spectral properties of visual pigments can be measured using a variety of methods such as microspectrophotometry, single-cell recordings, and electroretinograms. The relative ease with which physiological data on the visual pigments can be collected, coupled with the availability of molecular sequence data, makes opsins a useful system for studying the relationship between a protein's function and structure. A comparative approach has been used to identify changes at amino acid sites that have spectral tuning effects in New World monkeys (Neitz, Neitz, and Jacobs 1991; Shyue et al. 1998), fish (Yokoyama and Yokoyama 1990; Yokoyama et al. 1999), and other organisms (Chang et al. 1995). This approach makes use of a phylogeny upon which amino acids potentially involved in modulating the absorption spectrum of the visual pigment may be mapped. For instance, Chang et al. (1995) identified an amino acid substitution correlated with a blue shift in vertebrate and invertebrate opsins that was subsequently tested and experimentally confirmed using resonance Raman spectroscopy (Lin et al. 1998).

This study takes advantage of an available collection of lepidopteran long wavelength–sensitive opsin sequences (Chase, Bennett, and White 1997; Kitamoto et al. 1998; Briscoe 2000), and contributes the sequences of several naturally occurring visual pigment spectral variants from a pair of butterfly and moth species. λ_{max} values of these pigments have been characterized by others (Goldman, Barnes, and Goldsmith 1975; Langer, Haumann, and Meinecke 1979; Bernard 1983), whereas the sequences encoding the seven transmembrane domains are newly presented here. The data have allowed the examination of two sets of questions related to the description of the evolutionary history of lepidopteran opsins. First, what are the global patterns and processes governing long-wavelength (LW) opsin evolution following gene duplication and functional divergence? Second, can a comparative approach be taken for identify-

[1] Present address: Department of Cellular and Structural Biology, University of Colorado Health Sciences Center.

Key words: visual pigment, color vision, rhodopsin, photoreceptor, convergent evolution.

Address for correspondence and reprints: Adriana D. Briscoe, Department of Cellular and Structural Biology, University of Colorado Health Sciences Center, B-111, 4200 East 9th Avenue SOM Rm 4547, Denver, Colorado 80262. E-mail: adriana.briscoe@uchsc.edu.

Mol. Biol. Evol. 18(12):2270–2279. 2001

ing amino acid sites under positive Darwininan selection correlated with changes in opsin spectral function?

Materials and Methods

PCR, Cloning, Sequencing

Galleria mellonella and *Vanessa cardui* were obtained as pupae from Carolina Biological Supply Company (Burlington, NC) and reared in the lab until eclosion. Live adult specimens of *Precis coenia* were gifts from S. B. Carroll and adult specimens of *Spodoptera exigua* were gifts from T. J. Dennehy. Adult retinas were frozen in liquid nitrogen, homogenized, and total RNA from this material was extracted using Trizol (GIBCO-BRL), and resuspended in 100 μl diethyl pyrocarbonate-treated dH_2O. Single-stranded cDNA was synthesized from 10-μl total RNA using an adaptor primer and reverse transcriptase (GIBCO-BRL). Two sets of primers were then used in PCRs to obtain fragments of the long-wavelength opsins, 80 (5′-GAA CAR GCW AAR AAR ATG A-3′) and OPSRD (5′-CCR TAN ACR ATN GGR TTR TA-3′), which amplifies a short (242 nt) fragment of the gene, and 80-RACE2 (a primer complementary to the adaptor), which amplifies 3′-RACE products. Once gene-specific fragments had been cloned, species-specific reverse primers were designed to amplify a longer fragment (*Precis* RD, 5′-TAG TGT TAG CCG CAT CAG AGG-3′; *Vanessa* RD, 5′-ACA AGC AAA GGC ATA AGA AAT-3′; *Galleria* RD, 5′-TCA GAA GAG CGA AGA GAA GCG-3′; *Spodoptera* RD, 5′-GCT TGT GTT TGC CGC TTC AGA-3′), in combination with the forward primer LWFD2 (5′-CAY YTN ATH GAY CCN CAY TGG-3′). In all cases, the following PCR conditions were used: 94°C for 1 min then 35 cycles of 94°C for 30 s, 50°C for 1 min, 68°C for 1 min. To minimize the effects of *Taq* error, a proofreading DNA polymerase mix was used (Advantage *Taq* DNA polymerase, Clontech) in all PCR reactions, and overlapping pieces of the gene were amplified and cloned independently. PCR products were purified with the Prep-a-Gene kit (BioRad), and then cloned using the TOPO-TA cloning kit (Invitrogen). For each cloned PCR product, 2–6 clones were purified (QIAprep Spin Mini Preps, Qiagen) and sequenced with the ABI Big-Dye Terminator Reaction kit in forward and reverse directions and run on an ABI 377 Automated Sequencer. As a control for PCR contamination, the primer pair 80–OPSRD was also used to amplify an intron-containing region of genomic DNA from each of the four species. Sequences from these cloned genomic fragments were compared with sequences obtained from cDNA.

Phylogenetic Analyses

Opsins are a part of a large multi-gene family. To guard against the possibility that the cloned opsins belong to the UV or blue spectral classes, neighbor-joining (NJ) and maximum parsimony (MP) analyses of 43 sequences of spectrally diverse opsins were conducted. Their GenBank accession numbers are: *Apis mellifera* (Bee) (UV, AF004169; Blue, AF004168; LW, U26026), *Cambarellus ludovicianus* (AF003543), *Camponotos abdominalis* (UV, AF042788; LW, U32502), *Cataglyphis bombycinus* (UV, AF042787; LW, U32501), *Drosophila melanogaster* (Dmel) (Rh1, K02315; Rh2, M12896; Rh3, M17718; Rh4, M17730; Rh5, U67905; Rh6, Z86118), *Drosophila pseudoobscura* (Dpse) (Rh1, X56877; Rh2, X65878; Rh3, X65879; Rh4; X65880), *G. mellonella* (AF385330), *Heliconius sara* (OPS1, AF126753), *Hemigrapsus sanguineus* (Crab) (1, D50583; 2, D50584), *Limulus polyphemus* (lateral eye, L03781; ocelli, L03782), *Manduca sexta* (1, L78080; 2, L780801; 3, AD001674), *Papilio glaucus* (Papilio) (PglRh1, AF077189; PglRh2, AF077190; PglRh3, AF067080; PglRh4, AF077193; PglRh5, AF077191; PglRh6, AF077192), *Papilio xuthus* (Papilio) (PxRh1, AB007423; PxRh2, AB007424; PxRh3, AB007425), *P. coenia* (AF385332), *Procambarus clarkii* (Pclarkii) (S53494), *Schistocerca gregaria* (Locust) (1, X80071; 2, X80072), *Sphodromantis* spp. (Mantid) (X71665), *S. exigua* (AF385331), *V. cardui* (AF385333).

Amino acid sequences were aligned using Clustal W (Thompson, Higgins, and Gibson 1994), and then nucleotide sequences were aligned accordingly. Calculation of p-distance using the Nei–Gojobori method (Nei and Gojobori 1986) in MEGA (Kumar et al. 2000) suggested that the nucleotide sequences were too divergent to be reliably used in MP phylogenetic reconstruction (data not shown). Therefore, only amino acid sequences were used in phylogenetic analyses of this larger data set. A total of 282 amino acid sites were included in the MP and the NJ analyses. A stepmatrix of amino acid changes derived from a larger data set of G protein–coupled receptors (Rice 1994) was employed as a weighting scheme to account for unequal probabilities of amino acid change in the MP analysis. For the NJ analysis, total character distance (uncorrected p-distance) was used along with minimum evolution as the objective function. The reliability of the trees was tested by bootstrap analysis in PAUP* (Swofford 1998).

To establish the evolution of opsin genes within the long-wavelength lepidopteran opsin group, a smaller data set (13 sequences) was used. Tree topologies were estimated from nucleotide data in which all three positions were pooled using maximum likelihood algorithms under a variety of models of evolution (Yang 2000). The PAUP* computer program (Swofford 1998) was used for conducting tree searches and making estimates of the model parameters. Significant differences in the fit of the models to the data were tested using the likelihood ratio test (Hasegawa, Kishino, and Saitou 1991).

Statistical Analyses

The rate of molecular evolution has been observed sometimes to change following gene duplication. To explore this possibility, the two-cluster test of Takezaki, Rzhetsky, and Nei (1995) as implemented in PhylTest 2.0 (Kumar 1996) was used to test the constancy of the molecular clock between clusters consisting of the duplicated genes *Papilio* Rh1 and Rh3. As different parts of a protein may be under different levels of functional constraint, the data were divided into three categories—

Table 1
λ_{max} Values of Lepidopteran Long Wavelength Opsins

Species Sampled	GenBank No.	λ_{max} (nm)	References
Vanessa cardui	AF385333	530	Bernard 1983
Precis coenia	AF385332	510	G. Bernard, personal communication
Galleria mellonella.	AF385330	510	Goldman, Barnes, and Goldsmith 1975
Spodoptera exigua	AF385331	515	Langer, Haumann, and Meinecke 1979
Heliconius sara.	AF126753	550	Struwe 1972
Papilio xuthus PxRh1 . .	AB007423	520	Kitamoto et al. 1998; Arikawa et al. 1999
Papilio xuthus PxRh2 . .	AB007424	520	Kitamoto et al. 1998; Arikawa et al. 1999
Papilio xuthus PxRh3 . .	AB007425	575	Kitamoto et al. 1998; Arikawa et al. 1999
Manduca sexta	L78080	520	White et al. 1983

all, transmembrane (TM), and non-TM domains—to examine the relative contribution of each domain to any rate difference, if detected.

Domains under higher levels of constraint are expected to be less variable than domains in which this assumption is relaxed. As an additional measure of differences in the level of variability between domains, the ratio of the total number of sites in the TM domain to the total number of sites in the non-TM domains was compared to the ratio of the number of variable sites in the TM to the number of variable sites in the non-TM domains using Fisher's exact test.

Reconstruction of phylogenies can be affected by changes in the pattern of substitution (heterogeneity) (Galtier and Gouy 1995). To test the possibility that changes in the pattern of amino acid substitution have occurred following changes in the absorption spectra or following a duplication event, the Disparity Index (I_D) (Kumar and Gadagkar 2001) was calculated. The Disparity Index measures the extent of difference in evolutionary pattern between two sequences, and is a more powerful test than the χ^2 test (Kumar and Gadagkar 2001). This method allows the identification of sequences that are evolving under patterns that violate assumptions used in phylogenetic reconstruction and which probably should be removed from phylogenetic analyses.

Codon-Based Tests of Selection

A maximum likelihood approach (Nielsen and Yang 1998; Yang 2000) was used to test first for the rate heterogeneity between sites and then for the presence of amino acid sites and branches that may have $\omega > 1$, a signature of positive Darwinian selection. Multiple models of site-specific evolution were compared using the likelihood ratio test. Additionally, the value of ω was inspected to determine whether in the data a class of site under positive selection ($\omega > 1$) exists.

To test for the possibility of branch-specific $\omega > 1$, a model in which all branches in the tree are assigned the same ω (M0) was compared with a model in which all branches have a different ω value (free-ratio) (Yang 1998).

Ancestral State Reconstruction

The tree used for ancestral-state reconstructions was estimated using a maximum likelihood approach and a variety of models of evolution. The nucleotide sequences were then translated into amino acid sequences and imported into PAML (Yang 2000) where the aaml program, empirical + frequencies model, with the Jones, Taylor, and Thornton (1992) amino acid substitution rate matrix was used. Opsin amino acid ancestral states were also compared to reconstructions made in MacClade (Maddison and Maddison 1999). Patterns of amino acid substitution correlated with wavelength shifts in opsin absorption spectrum were then examined in two ways. First, amino acid sites undergoing substitution in the lepidopteran LW opsin tree were examined for homology to amino acid sites with known spectral tuning effects in vertebrate opsins (Neitz, Neitz, and Jacobs 1991; Asenjo, Rim, and Oprian 1994; Sun, Macke, and Nathans 1997). Secondly, the branches of the tree were examined for instances of convergent or parallel amino acid substitution. The CONVERGE program was used to test whether the observed number of parallel or convergent substitutions exceeded the expected (Zhang 1997; Zhang and Kumar 1997).

Results

Isolation of Four Lepidopteran LW Opsin cDNAs

PCR amplifications of cDNA using the primer pair 80–OPSRD (encoding the amino acids, QAKKMNV and YNPIVY) yielded single bands that were 242 bp in length in all species. PCR amplification of genomic DNA yielded single bands from *Precis* (341 bp), *Vanessa* (327 bp), *Galleria* (323 bp), and *Spodoptera* (242 bp). Each of the genomic PCR products contained an intron except for *Spodoptera* (*Precis*, 99 bp; *Vanessa*, 85 bp; *Galleria*, 81 bp). The coding regions of the genomic and cDNA sequences isolated from animals of the same species were identical at the nucleotide level, whereas the introns and coding regions were different between species. Gene-specific primers in combination with the reverse primer LWRD2 yielded longer fragments of the gene (*Vanessa* RD, 855 bp; *Precis* RD, 708 bp; *Galleria* RD, 690 bp; *Spodoptera* RD, 712 bp), which overlapped and matched identically with 80–OPSRD fragment amplified from each of these species (table 1 and fig. 1).

Opsin Gene Family Phylogeny

Invertebrate opsins are part of a multi-gene family in which peak absorption (λ_{max}) spans 345–575 nm

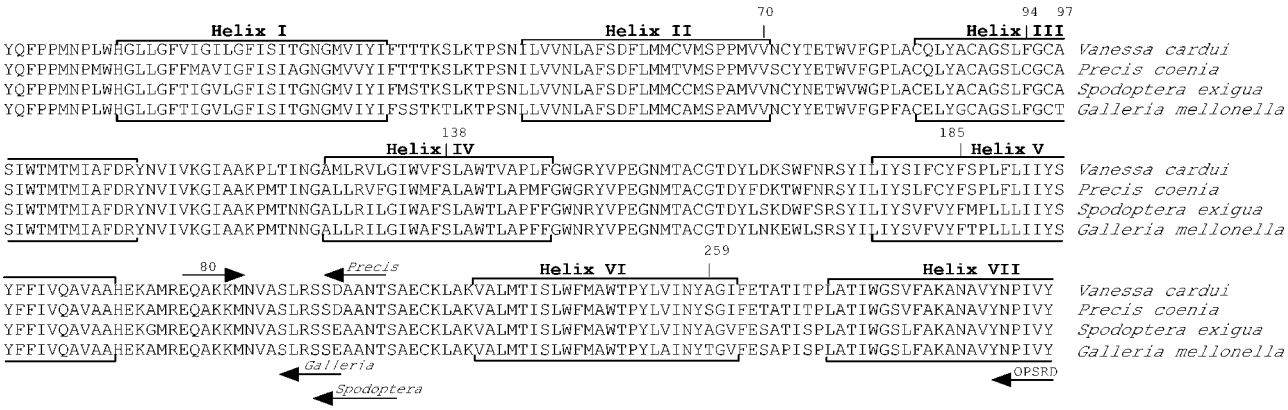

FIG. 1.—Alignment of cloned lepidopteran opsin sequences. Brackets indicate the seven transmembrane domains. Amino acid sites 138 and 185 are homologous to spectral tuning sites 180 in human and 229 in New World monkey cone pigments, respectively. Amino acid sites 70, 94, 97, and 259 are where substitutions have occurred more than once within the lepidopteran LW opsin clade, correlated with a spectral shift. Arrows indicate locations of primers used to isolate the DNA sequences. Forward primer LWFD2 is located just 5′ of amino acid site 1, and *Vanessa* RD primer is located in the 3′UTR (data not shown).

(Briscoe and Chittka 2001). Although a reverse primer specific to the LW opsins (OPSRD) was used to generate initial PCR products, it was important to rule out the possibility that the isolated opsins belonged to the UV- or blue-sensitive groups. The results of both NJ and MP analyses of 43 invertebrate opsins sequences are shown in figure 2. As indicated by the distribution of λ_{max} values across the tree, spectrally similar opsins cluster together phylogenetically (Briscoe 2000). The four new lepidopteran sequences fall squarely within a clade of long wavelength–sensitive (510–575 nm) opsins.

In addition, the new moth sequences from *Galleria* and *Spodoptera* cluster with a previously isolated moth (*Manduca*) sequence (Chase, Bennett, and White 1997), and the new butterfly sequences from *Precis* and *Vanessa* (both nymphalids) cluster with the nymphalid butterfly *Heliconius* sequence (Hsu et al. 2001). Neither the moth opsin cluster nor the nymphalid cluster shows close homology with the duplicated *Papilio* opsin genes (Rh1–4) in this initial analysis, although the relationships to other species within each of the clusters are consistent with their presumed biological relationships based upon morphological and molecular characters (Kristensen and Skalski 1998; Nylin et al. 2001).

The null hypothesis of homogeneity of amino acid substitution patterns between opsins could not be rejected using the Disparity Index test. Four statistically significant ($P < 0.05$) comparisons were identified out of 80, a number that corresponds exactly to the expected Type-I error rate for this number of tests (5% of 80 tests = 4). Therefore, all lepidopteran LW opsin sequences were retained for further analysis.

Duplicated LW Opsins Are Evolving at Different Rates of Amino Acid Substitution

The hypothesis of equal rates of amino acid substitution between the duplicated gene clusters, PglRh1–PxRh1 and PglRh3–PxRh3, was rejected at the 5% level when all sites were included (Z statistic = 2.8225) in the two-cluster test. This elevated rate of substitution along the PglRh3–PxRh3 branch seems to be primarily because of substitutions in the TM domains, as the Z statistic calculated with only the TM domain sites were significant at 5% level (Z = 2.7559), whereas the Z statistic calculated with the non-TM sites was not (Z = 0.93968).

When only the duplicated *Papilio* opsins were considered (PglRh1–4, PxRh1–3), the ratio of variable sites within the TM domain to non-TM domain (56:26) was significantly different ($P = 0.039$) from the ratio of the total number of TM domain sites to non-TM domain sites (160:122) using Fisher's exact test. This suggests that following gene duplication there has been a change in the relative level of variability in different domains, where there appears to be a higher level of variability in the TM domains relative to the non-TM domains. As a control for this scenario, the nonduplicated non-*Papilio* moth and butterfly opsins were similarly examined and displayed no significant difference in the variability between TM and non-TM domains ($P = 0.269$).

Evolutionary Relationships Within the Lepidopteran Long-Wavelength Opsin Family

Eight models of molecular evolution were used in maximum likelihood estimates of the lepidopteran LW opsin gene tree. The relative goodness-of-fit of the models to the data was tested using the likelihood ratio test (Hasegawa, Kishino and Saitou 1991; Yang 1995). The goodness-of-fit of the model to the data increased with increasing complexity of the model. The largest improvement in likelihood score of the recovered tree topology occurred between the HKY85 and HKY85 + Γ models (data not shown). HKY85 assumes a single substitution rate at all sites, whereas HKY85 + Γ assumes a site-to-site rate variation that follows the gamma distribution. This significant improvement in score between the two models is consistent with differences in %GC content among the first (42%), second (43%), and third (54%) positions of the data set, as well as with differences in the numbers of parsimoniously informative sites at first (69/282), second (44/282), and third (245/

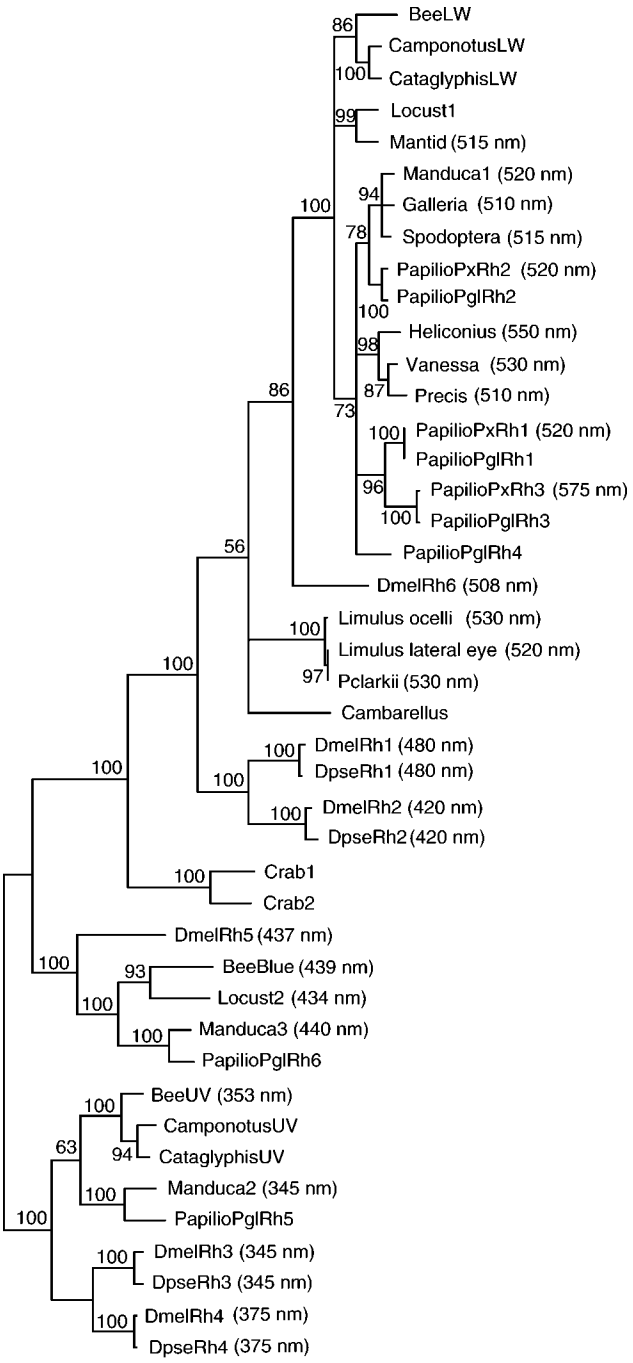

FIG. 2.—Phylogeny of arthropod opsins. The phylogeny is a bootstrap consensus tree based upon 1,000 replications of the neighbor-joining algorithm. Bootstrap values are given as a percentage and only bootstrap values >50% are shown. Branch lengths are drawn proportional to the total number of amino acid replacements that occur along that branch. This tree is identical to the maximum parsimony (MP) bootstrap consensus tree everywhere except that in the MP tree, the lepidopteran long-wavelength opsin clade collapses into a polytomy. Numbers in parentheses are the λ_{max} values of the visual pigments. For references see Briscoe (2000) and Briscoe and Chittka (2001). For list of taxa and explanation of abbreviations see *Materials and Methods.*

282) positions. Models that fit the data best (HKY85 + Γ, HKY85 + Γ + I, GTR + Γ, GTR + Γ + I) recovered an identical tree topology (fig. 3). Models that were a significantly worse fit to the data (JC69, F81, HKY85, and GTR) differed from this topology in their placement of Rh4. In these models, Rh4 clustered with other butterfly opsins as the most basal member of a monophyletic butterfly opsin clade. However, the difference between the two estimated topologies is not expected to have affected the handful of sites undergoing convergent evolution at branches near the tips of the tree (identified later), so the topology recovered using more realistic models of evolution (shown in fig. 3) was used in subsequent ancestral-state reconstruction.

Testing for Positive Selection

The likelihood ratio test comparing a neutral model (M1) with two rate categories for codon sites ($\omega = 0$ or 1) versus a model (M2) with an additional class of site estimated from the data rejected the hypothesis of two rate classes, conserved and neutral, across all sites (table 2). The same was true when M1 was compared with a model with three rate categories (M3) estimated according to a discrete distribution. Inspection of the three values of ω under M3 did not indicate the presence of a category of codon with $\omega > 1$ however, so none of the sites averaged over all branches display evidence for positive natural selection: a result that also held when models using a beta distribution were compared (M7 vs. M8). This is not entirely surprising because these sequences display a high level of divergence at the nucleotide level, which may have swamped out instances of positive selection that occurred early on in the evolution of new gene function. It is also possible that positive natural selection has only occurred along a few branches at a few sites, and averaging of ω over all branches may swamp out this signal.

To test this possibility, a free-ratio model was compared to a model (M0) that gives one ω value for all sites and all branches (table 2). The free-ratio model was in fact a better fit to the data than the M0 model and one branch in particular was identified as having a value of $\omega = 2.1044$ (fig. 3). However, the ratio of 11.2 inferred nonsynonymous substitutions per nonsynonymous site to 2 synonymous substitutions per synonymous site is not significantly different from the neutral expectation of 611.5 nonsynonymous to 234.5 sites (as estimated in PAML) (Fisher's exact test, $P = 0.260$). Therefore, the estimated ω value for this branch is not statistically significantly different from 1, and the hypothesis of neutral evolution cannot be rejected.

Ancestral-State Reconstructions

Similar shifts in visual pigment absorption spectrum maximum have evolved in vertebrates through convergent or parallel amino acid substitutions at specific sites in the opsin protein (e.g., Yokoyama and Radlwimmer 1999). To identify sites that may result in spectral shifts in the lepidopteran LW opsins, ancestral states were reconstructed using a maximum likelihood approach (data not shown). Most ancestral amino acid states were reconstructed with posterior probabilities between 90% and 100%: only 54 out of 1,128 sites had lower posterior probabilities. Comparison of ancestral

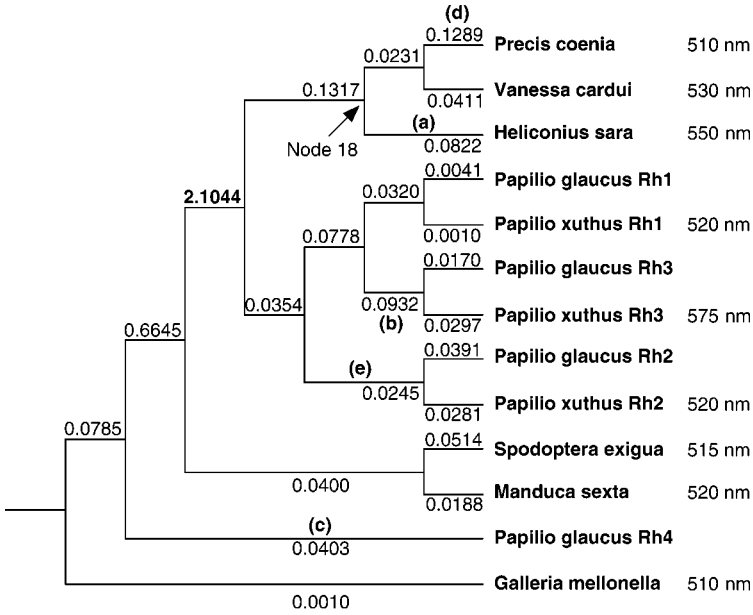

FIG. 3.—Maximum likelihood reconstruction of lepidopteran LW opsin phylogeny based upon analysis of nucleotide sequences (872 bp). The HKY85 + Γ, HKY85 + Γ + I, GTR + Γ, and GTR + Γ + I models all produced the topology shown. Parameter estimates under the HKY85 + Γ + I model: LnL = −6354.4364, ti/tv ratio = 2.290348, κ = 4.543528, proportion of invariable sites = 0.430770. Estimated value of gamma shape parameter = 0.898049. Branch-specific ω values are given as numbers. Small letters in bold and parentheses indicate branches examined in table 4 for convergent or parallel evolution.

states reconstructed in PAML using either a poisson correction or the empirical + base frequencies model, with those reconstructed in MacClade reveal a near-perfect degree of overlap. Only four sites (aa 41, 74, 143, 169) differed between the analyses. Maximum likelihood analyses have the added advantage of assigning differential probabilities to alternative ancestral states that are given equal weight in MP (Yang, Kumar, and Nei 1995), therefore these reconstructions were preferentially used to identify parallel or convergent changes.

Of particular interest are branches leading to opsins that have undergone large (20–45 nm) spectral shifts in absorption (λ_{max}) relative to ancestral opsins. Table 3 presents the reconstructed amino acid sites along branches undergoing large shifts in absorption spectrum. In the branch leading to the blue-shifted *Precis* opsin, there is an S to A substitution at aa 138. This substitution in human cone pigments (aa 180) has been shown to cause a 5-nm blue shift in site-directed mutagenesis experiments (Asenjo, Rim, and Oprian 1994). In a study of crayfish LW (520–530 nm) opsins, Crandall and Cronin (1997) found an F to C substitution at aa 102 (aa 94 in Lepidoptera) associated with a blue shift, a site that is also changing in parallel in the blue-shifted *Precis* lineage. Along the branch leading to the red-shifted *Heliconius* opsin sequence, an F to Y substitution has occurred at an amino acid site (aa 185) with known spectral tuning effects in New World monkeys (aa 229) (an F to I substitution at this site is associated with a 2-nm red shift) (Shyue et al. 1998). Parallel substitutions at aa 185 have also been correlated with blue shifts in crayfish LW opsins (Crandall and Cronin 1997).

In addition to sites in the *Precis* and *Heliconius* sequences that display parallel or convergent evolution with vertebrate or crayfish opsins, there are a handful of sites that have evolved in a similar fashion among different branches of the lepidopteran opsin clade. Specifically, amino acid sites 70, 94, and 97 have undergone parallel or convergent changes that are correlated with red shifts along the *Heliconius* (branch a, fig. 3) and *Papilio* Rh3 (branch b, fig. 3) branches (table 3). Using the program CONVERGE, the expected number of parallel and convergent changes given the branch lengths of the tree were compared with the observed. In both categories of change, the observed number of changes exceeded the expected at less than the 5% level of significance (table 4). All of these sites are located in the transmembrane domains of the opsin protein, and may have an effect on opsin spectral tuning (fig. 4). Besides these branches, along which large red shifts in absorption spectrum have occurred, there are three other pairs of branches with higher-than-expected numbers of parallel or convergent changes (table 4). Whether or not

Table 2
Likelihood Values Under Models of Heterogeneous ω Among Sites and Branches

Model	Neutral	Positive	$2\Delta\ell$[a]	df	*P*
Site variation					
M1 versus M2..	−6032.51	−5592.54	879.94	2	<0.0001
M1 versus M3..	−6032.51	−5577.37	910.28	4	<0.0001
M7 versus M8..	−5578.32	−5578.32	0.001	2	0.99
Lineage variation					
M0 versus Free-Ratio.....	−5695.60	−5653.06	85	23	<0.0001

NOTE.—Analyses were conducted using κ as a free parameter and the F3×4 model of equilibrium codon frequencies.

[a] Twice the log likelihood difference between the two models.

Table 3
Amino Acid Replacements Associated with Wavelength Shifts

AA Site	530 nm *Vanessa*	510 nm *Precis*	AA Site	530 nm Node 18	550 nm *Heliconius*	AA Site	530 nm *Papilio* Rh1	575 nm *Papilio* Rh3
9.....	L	M	1	Y	H	11	H	Y
17.....	V	F	26	I	V	17	V	T
18.....	I	M	33	I	V	20	V	C
19.....	G	A	41	S	T	22	G	A
20.....	I	V	62	C	F	23	F	I
21.....	L	I	63	V	M	24	I	T
27.....	T	A	65	S	A	26	L	I
33.....	I	V	**70**	**V**	**I**	27	T	S
62.....	C	T	74	Y	N	30	G	A
71.....	N	S	**94**	**F**	**Y**	41	T	N
74.....	T	Y	**97**	**A**	**V**	51	L	V
94.....	**F**	**C**	132	L	F	62	T	A
122.....	L	M	136	V	A	69	V	I
128.....	M	L	143	V	I	**70**	**V**	**I**
132.....	L	F	166	L	F	74	H	N
136.....	V	M	168	K	Q	83	A	F
138.....	*S*	*A*	170	W	F	86	L	I
143.....	V	L	171	F	S	89	A	C
146.....	L	M	178	I	L	90	A	G
166.....	L	F	182	F	A	**94**	**F**	**Y**
169.....	S	T	***185***	***F***	***Y***	**97**	**I**	**V**
181.....	I	L	186	S	A	104	M	A
259.....	**A**	**S**	263	E	K	123	T	S
			265	A	M	124	N	I
			270	L	I	136	V	L
			271	A	V	146	M	I
						160	A	V
						167	N	S
						171	F	L
						177	V	I
						181	I	V
						186	T	L
						189	A	G
						196	F	W
						225	E	D
						259	**T**	**A**
						261	I	V
						266	T	P
						270	L	V
						271	G	S

NOTE.—Amino acid replacements at the same site that occur more than once among lepidopteran opsins and are correlated with spectral shifts mapped onto the opsin phylogeny are bold. Amino acid replacements at sites with known spectral turning effects in human and New World monkey cone pigments are italic (see *Results*).

Table 4
Expected and Observed Numbers of Parallel and Convergent Amino Acid Changes

Branches	Parallel Changes	Convergent Changes	Sites
a and b	Expected = 0.337142 Observed = 2 (P = 0.045538)	Expected = 0.010772 Observed = 1 (P = 0.010714)	**70**, **94**, **97**
a and c	Expected = 0.3378 Observed = 2 (P = 0.045704)	Expected = 0.00932 Observed = 1 (P = 0.009328)	**62**, **186**, 265
a and d	Expected = 0.209522 Observed = 3 (P = 0.001311)	Expected = 0.000739 Observed = 0 (P = 1.0)	**33**, **132**, 166
a and e	Expected = 0.121657 Observed = 2 (P = 0.006826)	Expected = 0.001687 Observed = 0 (P = 1.0)	**33**, **185**

NOTE.—Only parallel or convergent substitutions reconstructed in PAML under all maximum likelihood models (poisson, empirical, empirical + frequencies) were included in the calculation. Expected number of changes calculated according to the method of Zhang and Kumar (1997) as implemented in the program CONVERGE. Exact probabilities of the observed number of changes given the expected are in parentheses. Branches tested are shown in figure 3. Amino acid sites located in the TM domains are in bold.

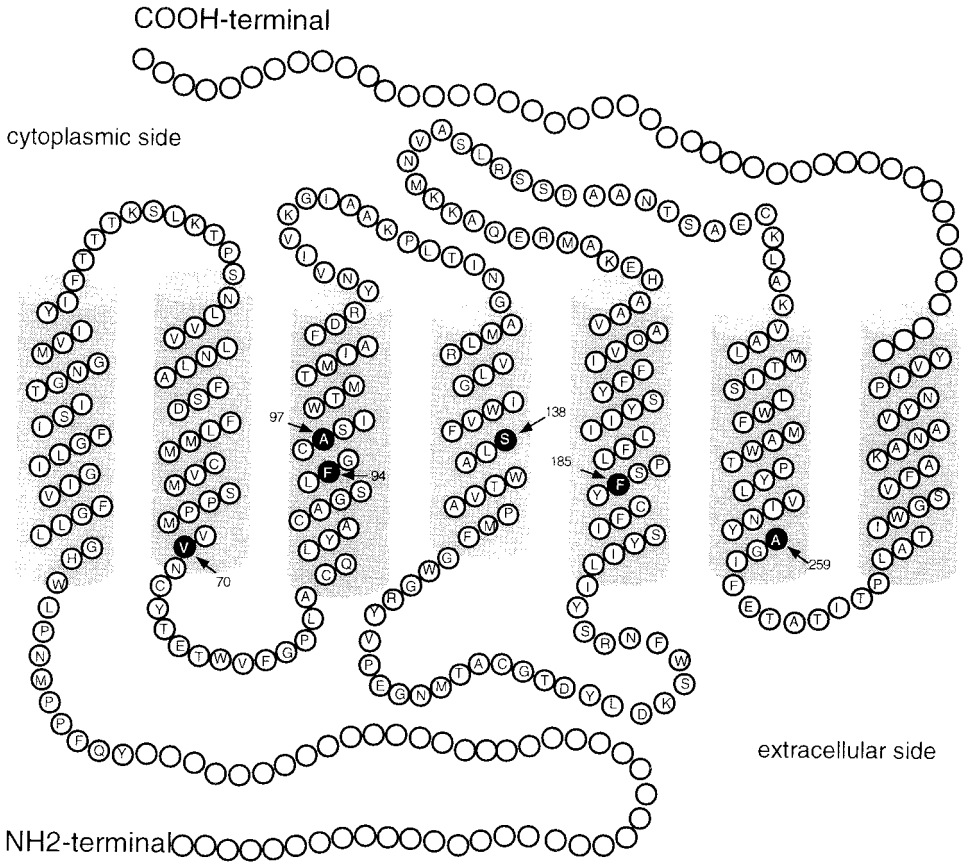

FIG. 4.—Topographical model of the *Vanessa cardui* LW opsin based upon a model of *Drosophila* Rh1 opsin from Britt et al. (1993). Sites of parallel or convergent amino acid evolution correlated with spectral shifts in absorption spectrum are indicated by black circles (see table 3).

they are correlated with changes in absorption spectrum is less clear because of missing (i.e., *Papilio* Rh4) or ambiguous functional data. In particular, the *Papilio* Rh1 and Rh2 opsins are likely to have slightly different λ_{max} values than the reported 520 nm because these measurements were taken from photoreceptor cells in which the two opsins are coexpressed (Kitamoto et al. 1998). Spectral sensitivity maxima of cells expressing two opsins have been shown to correspond to the summed sensitivity spectrum of the individual pigments (Feiler et al. 1992). More precise measurements of their λ_{max} values are needed to evaluate their potential adaptive significance, if any.

Discussion

The process by which genes acquire new functions is not well understood, but gene duplication is considered an important mechanism for functional divergence (Ohta 1993). Gene duplications are often followed by an accelerated rate of evolution (Ohta 1993), as observed in the *Papilio* Rh3 gene. The accelerated rate of evolution in the Rh3 gene, however, has not resulted in global changes in the patterns of amino acid substitution, as indicated by the results of the Disparity Index and Fisher's exact test, where the accelerated rate of evolution appears to have occurred only in the transmembrane domain of the protein. This finding suggests that the kinds of amino acid substitutions that are likely to result in spectral shifts are likely to be few in number, and predominantly found in the chromophore-binding pocket, consistent with models of opsin spectral tuning based upon site-directed mutagenesis experiments.

Using maximum likelihood–based tests of selection, an attempt was made to determine whether accelerated rates in the *Papilio* Rh3 gene were initially driven by positive Darwinian selection for functional divergence (Ohta 1993) or by relaxation of selective constraints. In the latter case, referred to as the Dykhuizen–Hartl effect (Zhang, Rosenberg, and Nei 1998), random fixation of neutral changes eventually leads to a novel function for one or both copies. The results of the branch-based test of selection, where the $\omega < 1$ for the *Papilio* Rh3 lineage, so far suggest a role for the latter model.

These results do not rule out the possibility that positive Darwinian selection has occurred at a handful of amino acid sites responsible for opsin spectral tuning, however. The *Papilio* Rh3 gene has in fact evolved a novel function relative to its ancestral gene, *Papilio* Rh2: it has evolved to be red-sensitive (575 nm) relative to its green-sensitive (~520 nm) ancestor. Evidence for positive Darwinian selection on this protein is derived

from the identification of several convergent and parallel amino acid substitutions (aa 70, 94, 97) that are correlated with red shifts along two independently evolved branches of the tree: the *Papilio* Rh3 (575 nm) and *Heliconius* (550 nm) lineages. At least one class of change, F to Y at aa 94, is of the category of amino acid substitution known to have spectral tuning effects in vertebrate opsins (i.e., the substitution of a hydroxyl-bearing Y for a nonhydroxyl bearing F). Indeed, this substitution, when mapped onto the three-dimensional crystal structure of bovine rhodopsin (Palczewski et al. 2000) is located in the third transmembrane domain of the protein (fig. 4), facing the chromophore-binding pocket. It would be worth testing the spectral tuning effects of this substitution in vitro.

Processes shaping the evolution of the paralogous *Papilio* Rh1 lineage may be a good example of what Force et al. (1999) refer to as the 'partitioning of ancestral [gene] functions rather than the evolution of new functions.' Consider the spatial expression pattern of Rh1 relative to its ancestral gene Rh2. Rh2 is expressed in three classes of photoreceptor cells in the *Papilio* retina: the R3–4 class, the R5–8 class, and the R9 class (Kitamoto et al. 1998). Rh1, the descendant gene, is expressed only in a subset of these photoreceptor cell classes, R3–4 and R9. Moving to a scale beyond individual photoreceptor cell classes within an ommatidium (the structural unit of the compound eye) to global opsin expression patterns across the eye, Rh2 is expressed both dorsally and ventrally in the R3–4 photoreceptor cells, whereas Rh1 expression is restricted ventrally (Kitamoto et al. 1998). Additionally, Rh1 and Rh2 are coexpressed in the R3–4 cells in the ventral part of the retina, suggesting that they may share overlapping regulatory elements. In the R9 cells, these duplicated genes seem to have nonoverlapping expression patterns.

Spectrally, Rh1 and Rh2 are very similar in absorption spectrum (~520 nm), though probably not identical (see earlier paragraphs of *Discussion*). Therefore, changes in the regulatory elements of Rh1 seem to have predominated following duplication, leading to the partitioning of ancestral Rh2 spatial expression patterns, rather than changes in the coding region of the gene which could lead to diversification of spectral function. In contrast, both partitioning of ancestral Rh2 spatial expression patterns and the evolution of novel function, i.e., red-sensitivity, appears to have occurred to the duplicated Rh3 gene. Rh3 is expressed in a subset of photoreceptor cells (R5–8) that also express Rh2 (Kitamoto et al. 1998). In some ommatidia Rh3 is coexpressed with Rh2 in the R5–8 cells, in other ommatidia only one of the two gene transcripts are expressed. It would appear that both complementary loss of gene subfunctionalizations and the acquisition of novel functions are mechanisms for the preservation of the *Papilio* opsin gene duplicates (Force et al. 1999).

Has the null hypothesis of neutral evolution been ruled out for any other part of the lepidopteran LW opsin clade? The observation of a parallel S to A change between vertebrate red and green cone pigments and the blue-shifted *Precis* (510 nm) opsin at an amino acid site in vertebrates (aa 180) with a known spectral tuning effect of 5 nm (Asenjo, Rim, and Oprian 1994), suggests that this site (aa 138 in *Precis*) may be under positive Darwinian selection in the *Precis* lineage. Similarly, another candidate site under positive Darwinian selection is the F to Y change from node 18 to the red-shifted *Heliconius* opsin at a site (aa 185) that in New World monkey cone pigments is associated with a 2-nm red shift (Shyue et al. 1998).

In conclusion, spectral diversification of the lepidopteran LW opsin clade (510–575 nm) appears to be driven by a handful of amino acid sites, some of which are shared with known spectral tuning sites in primates and fish, and others of which have so far only been identified in other invertebrate opsins, i.e., crayfish (Crandall and Cronin 1997). It will be interesting to see whether these same sites are involved in the spectral diversification of the UV and the blue insect opsins, but to do so will involve sampling a broader spectral range of pigments than are currently known.

Acknowledgments

I thank Sudhir Kumar, Steve Britt, and two anonymous reviewers for helpful comments on the manuscript, and Lisa Nagy and Allan Szeto for technical assistance. This work was supported by a Ford Foundation Postdoctoral Fellowship to A.B., and an N S F grant IBN 0082700 to S.B. and A.B.

LITERATURE CITED

ARIKAWA, K., D. G. W. SCHOLTEN, M. KINOSHITA, and D. G. STAVENGA. 1999. Tuning of photoreceptor spectral sensitivities by red and yellow pigments in the butterfly *Papilio xuthus*. Zool. Sci. **16**:17–24.

ASENJO, A. B., J. RIM, and D. D. OPRIAN. 1994. Molecular determinants of human red/green color discrimination. Neuron **12**:1131–1138.

BERNARD, G. D. 1983. Bleaching of rhabdoms in the eyes of intact butterflies. Science **219**:69–71.

BRISCOE, A. D. 2000. Six opsins from the butterfly *Papilio glaucus*: molecular phylogenetic evidence for paralogous origins of red-sensitive visual pigments in insects. J. Mol. Evol. **51**:110–121.

BRISCOE, A. D., and L. CHITTKA. 2001. The evolution of color vision in insects. Annu. Rev. Entomol. **46**:471–510.

BRITT, S. G., R. FEILER, K. KIRSCHFELD, and C. S. ZUKER. 1993. Spectral tuning of rhodopsin and metarhodopsin in vivo. Neuron **11**:29–39.

CHASE, M., R. R. BENNETT, and R. H. WHITE. 1997. Three opsin-encoding cDNAs from the compound eye of *Manduca sexta*. J. Exp. Biol. **200**:2469–2478.

CHANG, B. S. W., K. A. CRANDALL, J. P. CARULLI, and D. L. HARTL. 1995. Opsin phylogeny and evolution: a model for blue shifts in wavelength regulation. Mol. Phylogenet. Evol. **4**:31–44.

CRANDALL, K. A., and T. W. CRONIN. 1997. The molecular evolution of visual pigments of freshwater crayfishes (Decapoda: Cambaridae). J. Mol. Evol. **45**:524–535.

FEILER, R., R. BJORNSON, K. KIRSCHFELD, D. MISMER, G. M. RUBIN, D. P. SMITH, M. SOCOLICH, and C. S. ZUKER. 1992. Ectopic expression of ultraviolet rhodopsins in the blue photoreceptor cells of *Drosophila*: visual physiology and

photochemistry of transgenic animals. J. Neurosci. **12**: 3862–3868.

FORCE, A., M. LYNCH, F. PICKETT, A. AMORES, Y.-L. YAN, and J. POSTLETHWAIT. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151**: 1531–1545.

GALTIER, N., and M. GOUY. 1995. Inferring phylogenies from DNA sequences of unequal base compositions. Proc. Natl. Acad. Sci. USA **92**:11317–11321.

GOLDMAN, L. J., S. N. BARNES, and T. H. GOLDSMITH. 1975. Microspectrophotometry of rhodopsin and metarhodopsin in the moth *Galleria*. J. Gen. Physiol. **66**:383–404.

HASEGAWA, M., H. KISHINO, and N. SAITOU. 1991. On the maximum likelihood method in molecular phylogenetics. J. Mol. Evol. **32**:443–445.

HSU, R., A. D. BRISCOE, B. S. W. CHANG, and N. E. PIERCE. 2001. Molecular evolution of a long wavelength–sensitive opsin in mimetic *Heliconius* butterflies (Lepidoptera: Nymphalidae). Biol. J. Linn. Soc. **72**:435–449.

JONES, D. T., W. R. TAYLOR, and J. M. THORNTON. 1992. The rapid generation of mutation data matrices from protein sequences. Comp. Appl. Biosci. **8**:275–282.

KITAMOTO, J., K. SAKAMOTO, K. OZAKI, Y. MISHINA, and K. ARIKAWA. 1998. Two visual pigments in a single photoreceptor cell: identification and histological localization of three mRNAs encoding visual pigment opsins in the retina of the butterfly *Papilio xuthus*. J. Exp. Biol. **201**:1255–1261.

KRISTENSEN, N. P., and A. W. SKALSKI. 1998. Phylogeny and palaeontology. Pp. 7–25 *in* N. P. KRISTENSEN, ed. Lepidoptera, moths and butterflies, Vol. 1. Walter de Gruyter, Berlin.

KUMAR, S. 1996. Phyltest: a program for testing phylogenetic hypothesis. Version 2.0. Institute for Molecular Evolutionary Genetics and Department of Biology. The Pennsylvania State University, University Park, Pennsylvania.

KUMAR, S., and S. R. GADAGKAR. 2001. Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. Genetics **158**:1321–1327.

KUMAR, S., K. TAMURA, I. B. JAKOBSEN, and M. NEI. 2000. MEGA2: molecular evolutionary genetics analysis software. Arizona State University, Tempe, Arizona.

LANGER, H., B. HAUMANN, and C. C. MEINECKE. 1979. Tetrachromatic visual system in the moth *Spodoptera exempta* (Insecta: Noctuidae). J. Comp. Physiol. **129**:235–239.

LIN, S. W., G. G. KOCHENDOERFER, K. S. CARROLL, D. WANG, R. A. MATHIES, and T. P. SAKMAR. 1998. Mechanisms of spectral tuning in blue cone pigments. J. Biol. Chem. **273**: 24583–24591.

MADDISON, W. P., and D. R. MADDISON. 1999. MacClade: analysis of phylogeny and character evolution. Version 3.08. Sinauer Associates, Sunderland, Mass.

MERBS, S. L., and J. NATHANS. 1993. Role of hydroxyl-bearing amino acids in differentially tuning the absorption spectra of the human red and green cone pigments. Photochem. Photobiol. **58**:706–710.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonynous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

NEITZ, M., J. NEITZ, and G. H. JACOBS. 1991. Spectral tuning of pigments underlying red-green color vision. Science **252**: 971–974.

NIELSEN, R., and Z. H. YANG. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148**:929–936.

NYLIN, S., K. NYBLOM, F. RONQUIST, N. JANZ, J. BELICEK, and M. KÄLLERSJÖ. 2001. Phylogeny of *Polygonia*, *Nymphalis* and related butterflies (Lepidoptera: Nymphalidae): a total evidence analysis. Zool. J. Limn. Soc. Lond. **132**:441–468.

OHTA, T. 1993. Pattern of nucleotide substitution in growth hormone-prolactin gene family: a paradigm for evolution by gene duplication. Genetics **134**:1271–1276.

PALCZEWSKI, K., T. KUMASAKA, T. HORI et al. (12 co-authors). 2000. Crystal structure of rhodopsin: a G protein-coupled receptor. Science **289**:739–745.

RICE, K. 1994. The origin, evolution, and classification of G protein-coupled receptors. Doctoral dissertation, Harvard University.

SHYUE, S. K., S. BOISSINOT, H. SCHNEIDER et al. (13 co-authors). 1998. Molecular genetics of spectral tuning in New World monkey color vision. J. Mol. Evol. **46**:697–702.

SUN, H., J. P. MACKE, and J. NATHANS. 1997. Mechanisms of spectral tuning in the mouse green cone pigment. Proc. Natl. Acad. Sci. USA **94**:8860–8865.

STRUWE, G. 1972. Spectral sensitivity of the compound eye in butterflies (*Heliconius*). J. Comp. Physiol. **79**:191–196.

SWOFFORD, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0. Sinauer Associates, Sunderland, Mass.

TAKEZAKI, N., A. RZHETSKY, and M. NEI. 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. **12**:823–833.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

WHITE, R. H., P. K. BROWN, A. K. HURLEY, and R. R. BENNETT. 1983. Rhodopsins, retinula cell ultrastructure and receptor potentials in the developing pupal eye of the moth *Manduca sexta*. J. Comp. Physiol. **150**:153–163.

YANG, Z. 1995. Evaluation of several methods for estimating phylogenetic trees when substitution rates differ over nucleotide sites. J. Mol. Evol. **40**:689–697.

———. 1998. Likelihood ratio tests for detecting positive selection and applications to primate lysozyme evolution. Mol. Biol. Evol. **15**:568–573.

———. 2000. PAML: phylogenetic analysis by maximum likelihood. Version 3.0a. University College London, London, United Kingdom.

YANG, Z., S. KUMAR, and M. NEI. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics **141**:1641–1650.

YOKOYAMA, R., and S. YOKOYAMA. 1990. Convergent evolution of the red- and green-like visual pigment genes in fish, *Astyanax fasciatus*, and human. Proc. Natl. Acad. Sci. USA **87**:9315–9318.

YOKOYAMA, S., and F. B. RADLWIMMER. 1999. The molecular genetics of red and green color vision in mammals. Genetics **153**:919–932.

YOKOYAMA, S., H. ZHANG, F. B. RADLWIMMER, and N. S. BLOW. 1999. Adaptive evolution of color vision of the Comoran coelacanth (*Latimeria chalumnae*). Proc. Natl. Acad. Sci. USA **96**:6279–6284.

ZHANG, J. 1997. CAPE: convergent and parallel evolution at the amino acid level. Pennsylvania State University. University Park, Pennsylvania.

ZHANG, J., and S. KUMAR. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. Mol. Biol. Evol. **14**:527–536.

ZHANG, J., H. F. ROSENBERG, and M. NEI. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95**:3708–3713.

BRANDON GAUT, reviewing editor

Accepted August 27, 2001