2011-05-11

Characterization of Changes in Megalagrion Opsin Genes to Detect Signatures of Selection

Bhagya G. Janananda
University of Miami, bhagya@bio.miami.edu

UNIVERSITY OF MIAMI

CHARACTERIZATION OF CHANGES IN *MEGALAGRION* OPSIN GENES TO DETECT SIGNATURES OF SELECTION

By

Bhagya G. Janananda

A THESIS

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Master of Science

Coral Gables, Florida

May 2011

UNIVERSITY OF MIAMI

A thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

CHARACTERIZATION OF CHANGES IN *MEGALAGRION* OPSIN GENES TO DETECT SIGNATURES OF SELECTION

Bhagya G. Janananda

Approved:

_____	_____
Athula H. Wikramanayake, Ph.D.	Terri A. Scandura, Ph.D.
Professor of Biology	Dean of the Graduate School

_____
Barbara A. Whitlock, Ph.D.
Associate Professor of Biology

_____
Marguerite A. Butler, Ph.D.
Associate Professor of Biology
University of Hawaii at Manoa

JANANANDA, BHAGYA G. (M.S., Biology)
Characterization of Changes in *Megalagrion* Opsin Genes to Detect Signatures of Selection. (May 2011)

Abstract of a thesis at the University of Miami.

Thesis supervised by Professor Athula Wikramanayake
No. of pages in text. (50)

Megalagrion damselflies have radiated into new breeding habitats independently at least six times in the Hawaiian archipelago, and have evolved bright body coloration numerous times. We hypothesize that these radiations are correlated with specific changes in the opsin proteins. We isolated and characterized two opsin genes from nine different *Megalagrion* species. The opsin phylogeny is consistent with the phylogeny based on breeding habitat preference of *Megalagrion* species supporting the correlation between the evolutionary changes of vision and habitat shifts. dN/dS ratios of opsin sequences show that these genes are evolving under purifying selection, though some sites of the opsin genes might be evolving under positive selection. Two terrestrial-breeding *Megalagrion* species show higher rates of opsin gene evolution that are correlated with a rapid transformation in their breeding habitats from aquatic to terrestrial. These results support the hypothesis that opsin gene evolution has played a role in *Megalagrion* radiation in Hawaii.

ACKNOWLEDGMENTS

First and foremost I would like to thank my committee Dr Athula Wikramanayake, Barbara Whitlock and Marguerite Butler for their support and guidance throughout this project. I owe my most sincere gratitude to Professor Athula Wikramanayake for his time, ideas and encouragement as my advisor. I also thank all past and present members of the Wikramanayake lab for their support.

I would like to thank my outside committee member, Dr Marguerite Butler for providing primers and protocols and sample collecting. In regards to the help during field visits and sample collecting, I also thank Jeff Scales, David Preston and Roland Shroeder. My special thanks go to Marilou Sison-Mangus at UC Irvine for providing valuable information including protocols and primer sequences. Rebecca Duncan has been a source of friendship as well as good advice and collaboration on sequence analysis and Phylogenetic studies. I very much appreciate the efforts of Carl Sung, Naveen wijesena and Shalika Kumburegama in teaching me valuable laboratory techniques. The faculty, staff and friends at the Department of biology are warmly appreciated for their constant support during my time at the University of Miami.

Finally, I want to express my greatest appreciation for my parents, M.D. Janananda and Swarnalatha Weerarathne and my brother Hasitha Janananda for their love, understanding, patience and endless support. And most of all for my loving husband, Naveen and son, Mevan for their patience and support at all times. Thank you.

TABLE OF CONTENTS

Page

LIST OF FIGURES v

LIST OF TABLES vi

Chapter 1: INTRODUCTION 1

Chapter 2: METHOD 6

1. Isolating the UV and LW opsin genes from *Megalagrion* species and reconstruction of phylogeny…… 6
2. Deducing the secondary structure and tracing the amino acid substitutions important in wavelength discrimination…… 10
3. Test for the signatures of selection..... 11

Rate of molecular evolution

Chapter 3: RESULTS …..…… 15

1. Isolating the UV and LW opsin genes from *Megalagrion* species and reconstruction of phylogeny…… 15
2. Deducing the secondary structure and tracing the amino acid substitutions important in wavelength discrimination…… 17
3. Test for the signatures of selection…….. 19
4. Rate of molecular evolution…… 20

Chapter 4: DISCUSSION 21

REFERENCES..... 26

FIGURES LEGENDS…….. 30

LIST OF TABLES…… 32

FIGURES …… 34

TABLES …….. 43

LIST OF FIGURES

Page

1: Schematic diagram of a rhodopsin molecule showing the opsin moiety and the chromophore (11-cis-retinal) embedded within the transmembrane domains …….. 34

2.1: Bayesian tree of *Megalagrion* damselfly species relationships reconstructed using the mixed-model for amino acid sequences in MrBayes 3. 35

2.2: Bayesian tree of *Megalagrion* damselfly species reconstructed using the HKY model for nucleotide sequences in MrBayes 3.1……. 36

3.1: CLUSTAL W 2.0.12 multiple sequence alignment of the deduced amino acids of the UV opsin genes of nine species of Megalagrions and other closely related insect UV opsins..... 37

3.2: Nucleotide and deduced amino acid sequence of *Megalagrion* UV opsin…….. 39

3.3: A schematic diagram of the *Megalagrion* UV opsin protein designed using Adobe Illustrator…… 40

3.4: A simplified phylogeny showing the four conserved amino acids mapped on to phylogeny constructed by Chang et al., 1995…… 41

5.1: An illustration showing the comparison of data under the two-cluster relative-rate test as implemented in PHYLTEST software (Kumar 1996)…….. 42

LIST OF TABLES

Page

2.1 UV and LW gene specific and degenerate primers which were used in PCR…….. 43

3.1 Conserved amino acid residues that are important for specific wavelength detection…… 44

4.1 Parameters in Site Models…… 45

4.2 Maximum Likelihood pair wise comparison of dN/dS ratios of the UV opsin sequences of nine different *Megalagrion* species. Analysis was performed using a codon based substitution model of codeml, PAML (Nei & Gojobori 1986)……... 46

4.3 Parameters and Log-Likelihood Values in the Site Models of Variable ω Ratios among sites…….. 46

4.4 Likelihood Ratio Statistics……... 47

5.1 Cluster assignments in each analysis of relative rate test……. 47

5.2 Comparison of the rates of evolution among different *Megalagrion* species…….. 48

5.3 Comparison of the rates of evolution in between UV and LW sequences of different *Megalagrion* species…… 49

5.4 Comparison of the rates of evolution among different domains of the UV opsin protein of different *Megalagrion* species…….. 50

Chapter 1

INTRODUCTION

Different visual systems, adapted to different types of vision such as color vision, night and diurnal vision, polarized and UV vision, have been studied for many years in a number of different fields from biology to physics. Understanding how these different visual systems function was made possible by studying the absorbance spectra of different types of opsins. These studies have indicated that the variation in the absorbance spectra is primarily conferred by variation in the amino acid sequence of the visual pigments (Briscoe, 1998).

Rhodopsins or visual pigments, which are the basic functional units of visual systems consist of a protein moiety (an opsin) and a non-protein moiety, the chromophore retinal (a vitamin A-based retinaldehyde chromophore called 11-cis-retinal) where the opsin protein is covalently bound to the chromophore. (Figure1, Terakita, 2005).The opsins are transmembrane proteins that function as light sensors. These molecules are G-protein coupled receptors which mediate the activation of a downstream cascade (Towner, and Giirtnerb, 1994). Even though molecular evolution of vertebrate opsins is well studied, our knowledge of the evolution of many invertebrate opsins and visual systems is far from complete.

The signal transduction cascade leading to light sensing is initiated when a visual pigment absorbs a photon (Stavenga, 2006). Studies have shown that the photoreceptors of the retina function as the photon detectors that capture and transduce the photons

received from the external environment. The visual pigments which are embedded in the membranes of photoreceptor cells get excited by the light energy, which then causes phototransduction that triggers a chain of molecular processes culminating in a neural signal (Stavenga, 2006). Therefore visual pigments act as a visual information interface between the environment and the brain, and convert information on light intensity and spectral composition. Even minor changes in the amino acid sequence of the protein moiety can produce visual pigments with different peak sensitivities (Thompson, 1995).

Various animal models have been used to study visual system adaptations. Among these models, insect visual systems have been in the forefront for a number of reasons. Insects are a large and complex group of animals that show immense morphological, physiological and ecological diversity (Friedrich et al., 2006). A variety of diverse visual systems within insects have been the subject of numerous studies and most recent studies have been aimed at understanding the origin of visual system structures in arthropods (Harzsch et al., 2006). Molecular genetic investigations in *Drosophila* resulted in some of the most important advancements that have led to a better understanding of the development and structural organization of the insect visual system (Friedrich et al., 2006).

Among terrestrial animals, only vertebrates and arthropods have color vision or the ability of wavelength-discrimination (Koyanagi et al., 2008). Multiple opsins that code for the protein moiety of visual pigments sensitive to different wavelengths of light ranging from about 360 to 630 nm are required for color vision (Koyanagi et al., 2008). Insect photoreceptors belonging to three spectral classes have been identified and these include short wavelength-sensitive (sensitive to light less than 400 nm, eg; UV), medium

wavelength-sensitive (400 – 500 nm, eg; blue), and long wavelength-sensitive (>500nm, eg; green) photoreceptors (Spaethe, and Briscoe, 2004).

Studies related to insect vision, namely studies on spectral tuning, molecular characterizations and site-directed mutagenesis have been performed on opsin proteins found in many insect groups. Fruit flies, moths and butterflies, mantids and bees are some of the widely studied groups (Taylor et al., 2005). However the order Odonata, which includes dragonflies and damselflies, have been overlooked in studies related to vision. This study is focused on the characterization of the visual pigments of one such damselfly genus, *Megalagrion.*

Megalagrion **damselflies as a model system to study molecular evolution of color vision.**

The damselflies on which this study is focused on belong to the genus *Megalagrion* and are endemic to the Hawaiian archipelago. The genus *Megalagrion* consists of 23 described species that occupy the seven main Hawaiian Islands (Jordan et al., 2003). Previous studies have shown that the *Megalagrion* species are descended from a single ancestor that colonized the Hawaiian archipelago about 5.1 million years ago (Polhemus, 1996). A phylogeny based on 23 morphological and ecological characters implies an ecological progression of the *Megalagrion* genus from ancestral breeding sites in ponds or slow stream pools to breeding on seeps. Local clock analysis has shown that the species have radiated into fast streams and plant leaf axils on Kauai simultaneously and about 1.5 million years later another group moved simultaneously to seeps and terrestrial habitats on Oahu (Polhemus, 1997). Previous studies on Megalagrions show two speciation patterns. While some species colonized newly formed volcanic islands, other

species have rapidly radiated into a variety of specialized habitats within islands (Polhemus, 1997). At the same time, *Megalagrions* have evolved bright body coloration numerous times (Polhemus, 1997). The different habitats that these damselflies occupy differ in both physical and biological parameters such as temperature, light, elevation, food availability, and predatory pressure. Although *Megalagrion* species are closely related, they show extensive inter-specific color variation and striking sexual dichromatism (Jordan et al., 2003). In order to discriminate between these specific color variations, it is likely that their visual systems have been subjected to specific adaptations during evolution. All these characteristics make the endemic Hawaiian genus *Megalagrion* an excellent group for asking questions on natural selection and evolution of critical proteins involved in color vision.

Studies related to vision have shown a correlation between the photic environment and the sensitivity of the visual system (Spady et al., 2005). Species that depend on visual cues to find mates, food, and suitable habitats are expected to have visual system adaptations that have optimized their vision for the specific photic environment they live in (Lythgoe, 1979). *Megalagrions* are visually oriented animals that have acquired color vision and the ability to discriminate between different wavelengths of light. These make their visual systems an ideal model for addressing a number of different evolutionary questions. It is important, therefore, to characterize the genes that are important for vision in these insects in order to gain insights in to the selective pressures and constraints that drive visual system evolution in these insects.

In this study, I have sequenced one short wavelength opsin and part of a long wavelength detecting opsin gene of different *Megalagrion* spp. These sequences were

used to reconstruct a *Megalagrion* phylogeny and the amino acid changes were traced along the branches. Also, the sequence information was used to study different aspects of evolution of these gene sequences, namely the types of selection acting on these opsins and their rates of evolution.

Chapter 2

METHOD

1. Isolating the UV and LW opsin genes from *Megalagrion* species and reconstruction of phylogeny.

Isolating the UV and LW opsin genes from *Megalagrion* spp

Collecting specimens

Specimens of nine different species of *Megalagrion* damselflies found in different habitat types were collected. All specimens were collected using hand nets. These species include *M. calliphya, M. hawaiiense, M. heterogamias, M. nigrohamatum nigrolineatum, M. oahuense, M. oresitrophum, M. orobates, M. vagabundum* and *M. xanthomelas***.** Six out of the nine species are endemic either to Kauai or Oahu. *M. calliphya, M. hawaiiense* and *M. xanthomelas* are distributed in more than two islands. Opsin genes of some rare species (*hawaiiense, oahuense* and *calliphya)* were sequenced using specimens that were previously collected and preserved. The specimens were identified with the help of a field identification guide of Hawaiian Damselflies by Polhemus and Asquith (1997). The collected specimens were preserved in individual vials filled with RNA-later (Ambion, Austin, TX) solution and stored at -80ºC prior to RNA/DNA extractions.

RNA extraction, cDNA synthesis and PCR

Total mRNA was isolated from the eyes of different species of damselflies stored at -80 ºC. The eyes were dissected and weighed quickly to avoid sample thawing. Tissues

were homogenized with TRI reagent (Applied Biosystems) using a Teflon coated pestle and total RNA was extracted using the TRI reagent method **(**Chomczynski, 1993)**.** The extractions were done under RNAse-free conditions and on ice. Extracted total RNA was used to synthesize cDNA using random primers. A pair of primers for the 5' and 3' untranslated regions (UTR) of the *Megalagrion* UV opsin was used to amplify the complete UV opsin gene using the polymerase chain reaction (PCR). These primers (UV1F and UV1R) were designed for UTR sequences obtained using RACE (rapid amplification of cDNA ends). Primer UV1 F spans the 5' UTR and the reverse primer UV1R spans the 3' UTR of the UV opsin gene. The PCR program was set for 35 cycles of three steps each: (1) 30 sec at 94 ^{0}C (denaturing), (2) 1 min at 50^{0}C (annealing), (3) 30 sec at 72 ^{0}C (extension) and 2min at 72 ^{0}C (Elongation). This primer pair amplified a DNA fragment of approximately 1400bp. Then the PCR product was cleaned using EXOSAP-IT PCR clean up kit (GE Healthcare) and sent for sequencing with the primers that were used to amplify the fragment.

Genomic DNA extraction and PCR

Genomic DNA was extracted from thoracic and abdominal muscles using a standard phenol/chloroform method (Butterfly DNA extraction protocol was provided by Marilou Sison-Mangus, University of California, Irvine). Three sets of primers were used to amplify the complete UV gene using genomic DNA. The first primer set amplified a fragment of approximately 780 bp. The forward primer UVM1F was designed to the 5' UTR region and the reverse primer (UVM1R) was designed to the middle of the UV gene. The second set of primers UVM2F and UVM2R amplified a fragment of about 780

bp from the middle of the UV opsin gene. The reverse primer (UVM3R) of the third primer pair sits in the 3'UTR region of the opsin gene and this primer pair amplified a fragment of about 720 bp. Touchdown PCR was carried out using the three different primer sets. A set of gene specific LW primers (LW1F and LW1R) was used to isolate a 383bp fragment from the 3' end of the opsin gene and a set of degenerate primers (LW4F and LW4R) was used to amplify a fragment from the 5' end. All the primer sequences are given in table 2.1.

In order to test for hybridization, sample mix-up and possible contaminations, two individuals were sequenced from some species and the nucleotide sequences were aligned. If the sequences were identical, one of the sequences was used for the analysis. However, only one individual was sequenced from the rare species. These species will be re-sequenced if the results are found to be inconsistent.

Phylogenetic analysis

BLAST searches were carried out to identify each sequence and to further confirm that the sequenced opsins belong to the UV-clade and the LW-clade. Phylogenetic analyses using maximum likelihood criteria were carried out using MrBayes software (Huelsenbeck, and Ronquist, 2001). Apart from that, phylogenetic trees were constructed with PAUP (Swofford, 2003) using maximum parsimony and distance methods to obtain the tree files which were used in analysis using PAML (data not shown). Multiple sequence alignments for analyses using both MrBayes and PAUP including different types of opsin sequences of diverse insect taxa, other invertebrates and some vertebrates obtained from GenBank data base were generated using and Clustal X program. These

alignments and Phylogenetic trees were then used to determine the orthology of the *Megalagrion* opsins and to trace the important amino acid residues along the *Megalagrion* lineage.

Bayesian analysis

Three types of molecular phylogenies of opsin genes were constructed using bayesian methods as implemented in MrBayes 3.1. Three different alignment files were generated using Clustal X (Larkin et al., 2007), one with amino acid sequences (only the open reading frame) of closely related insect opsins along with the *Megalagrion* sequences and the other with nucleotide sequences of the open reading frames (ORFs) of the same species. The third included 43 opsin sequences (amino acid sequences of the ORFs) of diverse taxa, both vertebrate and invertebrate. The first and the second alignment files were used to construct phylogenies which would confirm the Phylogenetic placement of *Megalagrion* UV and LW opsins in the insect opsin tree. The phylogeny constructed using the third alignment file of 43 opsin sequences was used to trace the amino acid replacements along the lineages of the opsin phylogeny. For convenience the trees generated using the first second and third sequence files are named Tree-1, Tree-2 and Tree-3 respectively. Tree-1 and Tree-3 were generated using the mixed-model for amino acid sequences in MrBayes 3.1. The mixed model option allows the MCMC (Markov Chain Monte Carlo) sampler to explore all the fixed rate models implemented in Mr.Bayes by regularly proposing new models. When the MCMC procedure has converged, each model will contribute to the results in proportion to its posterior probability. The analysis involved a run of one million generations, sampled every 100

generations. The summary consensus tree was generated in MrBayes using the last 7500 trees and the posterior probabilities were calculated for this consensus tree.

Tree-2 was constructed using the HKY nucleotide model (Hasegawa, Kishino and Yano, 1985) as implemented in MrBayes 3.1, which has different rates for transitions and transversions. This analysis also involved a run of one million generations, sampled every 100 generations. The summary consensus tree was generated in MrBayes using the last 7500 trees and the posterior probabilities were calculated for this consensus tree.
Only Tree-1(Figure 2.1) and Tree-2 (Figure 2.2) are shown under results and tree-3 (data not shown) was used to trace the amino acid replacements.

2. Deducing the secondary structure and tracing the amino acid substitutions important in wavelength discrimination.

Comparative sequence analyses were carried out using the *Megalagrion* opsin sequences and different opsin sequences of other taxa to identify the conserved amino acid residues, and to determine the importance of these sites for opsin function and wavelength discrimination. First, the amino acid sequences of 43 opsin sequences, including both vertebrates and invertebrates were included along with the *Megalagrion* opsin sequences in the data set. Then multiple alignments were generated using ClustalX (Larkin et al., 2007) software programs. Then, the phylogeny reconstruction was carried out using Bayesian analysis as implemented in MrBayes 3.1.The amino acid changes at specific positions were determined based on the detailed amino acid alignment and the phylogeny used in Chang et al., (1995). Multiple sequence alignment generated using

Clustal W was used to identify the boundaries of putative transmembrane domains of the UV opsin G-protein coupled receptors.

A schematic diagram of the *Megalagrion* UV opsin protein was designed using Adobe Illustrator (Figure 3.3). Important amino acid residues and non-synonymous substitutions among nine *Megalagrion* species were indicated in the diagram.

3. Test for the signatures of selection.

***codeml* , PAML**

Maximum-likelihood models of codon substitution as implemented in *codeml* of PAML 4 (Yang 2007) were used to detect the type of selection acting on the *Megalagrion* opsin genes. Calculating the dN/dS ratio (ω) of both UV and Long wavelength opsins will indicate the selective pressures acting on these proteins. This model applies one ω ratio for the whole data set. The ω ratio is a measure of natural selection acting on the protein. Values of ω can be $\omega < 1$ (negative purifying selection), $\omega = 1$ (neutral evolution) and $\omega > 1$ (positive selection). Since the ω values are averaged over all sites or all lineages, ω is very rarely >1, indicating positive selection. Analyses focused on site or branch-site models which are widely used to detect positive selection which is acting on certain lineages or sites of a protein (Yang 2007).

Site models

This method uses several different site-specific models which allow the ω ratio to vary among sites. In this method, selective pressure varies among different sites but the site specific pattern is identical across all lineages (Yang et al., 2000). A codon or an amino acid is referred to as a site (Yang et al., 2000). *codeml* provides several site models

using the variable **Nssites**. The model is set as 0 (model=0) in the control file. All 6 different site models (M0, M1, M2, M3, M7 and M8) were used for the analysis by changing the variable **Nssites** in the control file, while the same sequence file and the tree file was used for all the analyses. The parameters used in each site model are summarized in Table 2. The selection types, purifying selection, neutral evolution and positive selection are denoted by ω values, $\omega < 1$, $\omega = 1$, and $\omega > 1$ respectively. For this analysis, the site models as implemented in *codeml* were used to detect the type of selection acting on UV opsin genes within the *Megalagrion* clade of the phylogenetic tree.

Three different files, the sequence data file, the tree file and the control file were used to run the analysis in both types of analysis. The sequence data file was converted to phylip format using MacClade (Maddison and Maddison, 2003) and the tree file was generated using PAUP.

After running different site models, a Likelihood Ratio Test (LRT) was carried out by comparing the likelihood ratio test statistic to critical values of the chi-square values with the determined df (degrees of freedom). This will give an idea of whether more complex models provide a significantly better fit to the data set than more simple models (Yang et al., 2000). Likewise three LRTs were calculated by comparing the M0 (one ratio) and M3 (discrete), M1a (neutral) and M2a (selection) and M7 (beta) and M8 (beta & w). The results are summarized in Table 4.3.

4. Rate of molecular evolution

Relative Rate Test

The relative rate test was carried out in order to compare the differences in the rates of evolution among different *Megalagrion* species and between different domains of the opsin proteins. The PHYLTEST software (Kumar 1996) was used to carry out the two-cluster relative-rate test of (Tajima, 1993), illustrated in Figure 5.1. First, the three clusters were specified based on the comparison (Table 5.1). Cluster A and B are the in-groups to be compared when an out group C is provided. PHYLTEST sets the equal rate of evolution or the consistency of the rate of evolution as the null hypothesis. PHYLTEST then simply tests for the acceptance or the rejection of the null hypothesis under the constancy of molecular clock. Constancy of the molecular clock between two lineages (A and B) is tested when an out group lineage (C) is given (Figure 5.1)

To examine the rate differences among opsins of different *Megalagrions*, the UV and LW opsin sequences of the *Megalagrion* species (Clusters A and B) were compared with the UV opsin sequence of *Apis* as the closely related outgroup (Cluster C). The rate of evolution of transmembrane and non-transmembrane domains of the UV opsin protein was also compared with the equal rates of molecular evolution as the null hypothesis. Since different regions of a protein can be under different functional constraints, the opsin was divided into three sections as transmembrane (TM), non-transmembrane (NTM) and complete (the entire open reading frame) to examine the rates of amino acid substitution. During this analysis, the transmembrane (TM) regions and the non-transmembrane (NTM) (clusters A and B) regions were compared with the complete

opsin protein sequence which assigns as cluster C (Table 5.1). PHYLTEST also allows the selection of a distance estimation method for both nucleotide and protein data. These methods correct for multiple substitutions by taking into account transition/transversion biases and varying substitution rates among sites. Of these methods, Jukes-Cantor distance (Jukes and Cantor 1969) method, which can specify the number of nucleotide substitutions per site, was selected for the *Megalagrion* nucleotide data set. For the protein data set, the Poisson correction method under the Relative rate test was selected.

Chapter 3

RESULTS

1. Isolating the UV and LW opsin genes from *Megalagrion* species and reconstruction of phylogeny.

Isolation of complete ORF of *Megalagrion* UV opsins

The sequences obtained using cDNA and UTR primers (UV1F and UV1R) contain 1363 base pairs (bp) with a single open reading frame coding for an opsin with 387 amino acids. The three sets of gene specific primers were used to sequence three overlapping DNA fragments of different sizes. Those fragments were then used to deduce the complete open reading frame of the *Megalagrion* UV opsin. Sequence comparisons with other insect UV opsin sequences along with the DNA structure predicting programs resulted in the identification of a gene sequence consisting of 1161 bp and a single open reading frame coding for an opsin with 387 amino acids. The deduced amino acid sequence of the *Megalagrion* UV opsin showed a high degree of sequence similarity to several other previously identified invertebrate opsins (Figure 3.1). Similarities at key amino acid positions shown in Table 3.1 along with the high degree of sequence similarity with other well characterized UV opsins (Figure 3.1), support the classification of isolated *Megalgrion* sequences as UV opsins which absorb light at a wavelength of approximately at 345-375nm.

Partial isolation of a Long Wavelength (LW) opsin of *Megalagrion* spp

A PCR fragment of 383 base pairs from the 5'region of the gene was amplified with the gene specific primers LW1F and LW1R (Table 1) for five different *Megalagrion* species. PCR amplification of another fragment of ~ 445 base pairs from the 3' end was also carried out using only 4 different species. Therefore only the translated amino acid sequences from the 5' region of the LW opsin of 5 different *Megalagrion* species were included in the Bayesian and Relative rate test analysis. High sequence similarity with other insect LW opsins and the Bayesian phylogeny confirmed the identity of these sequences as LW opsins.

Bayesian Phylogeny

Phylogenies constructed using both amino acid sequence data (Figure 2.1) and nucleotide sequence data (Figure 2.2) show monophyletic clades for both *Megalagrion* UV and LW opsin sequences. The placement of *Megalagrion* opsin sequences within the insect UV opsin clade and the insect LW opsin clade respectively, suggests that their absorption spectra are in the UV and LW ranges. These results along with Blast search results confirm the identity of *Megalagrion* sequences as UV and LW detecting opsins. *Megalagrion* UV opsins are more closely related to the *Apis cerana* UV-sensitive opsins while the *Megalagrion* LW cluster with *Drosophila* rhodopsins as shown in the Phylogenetic trees.

In the phylogeny reconstruction using UV opsin sequences, *M. heterogamias, M. oresitrophum*, *M. calliphya,,M. xanthomelas*, *M. orobates* and *M. nigrohamatum* cluster together in one clade (Figure 2.1 and 2.2). Based on Jorden et al. (2003), all the above

mentioned species use ponds, pools and streams as their breeding grounds while *M. hawaiiense, M. oahuense* and *M. vagabundum* that forms another clade in the UV opsin phylogeny use more terrestrial breeding habitats such as seeps and leaf axils. Therefore, the opsin phylogeny constructed in this study using molecular data is consistent with the phylogeny constructed based on *Megalagrion* damselfly breeding habitat usage.

2. Deducing the secondary structure and tracing the amino acid substitutions important in wavelength discrimination.

The deduced amino acid sequences of both UV and LW opsins show the various conserved characteristics of opsins in general and of arthropod opsins in particular. Sequence similarities were considered in assigning the putative transmembrane domains of these G-protein coupled receptors. Based on the sequence similarities, I have isolated the complete UV opsin sequence with the seven transmembrane domains that characterize opsins and other G-protein coupled receptors. Only UV opsin sequences are used for structure prediction.

Megalagrion UV opsin sequences share a number of structural features with other insect visual pigments and G protein-coupled receptors. These include conserved amino acid residues that are important in determining the function of the rhodopsin. Briscoe (1999) identified seven amino acid motifs that are important for a functional opsin protein and that are conserved across all known insect opsins. All seven of these conserved motifs are found in the *Megalagrion* UV opsins (Figure 3.1).

Lysine at amino acid position 296 of bovine rhodopsin in the transmembrane (TM) domain 7 is found in *Megalagrion* UV opsin at position 337. The chromophore is bound

by a protonated Schiff base linkage at this position (Briscoe, 1999). Cysteines at the positions 139 and 217 which form the disulfide bridge connecting TM region 3 with the second cytostolic loop are conserved in Megalagrions. Leucine at position 98 and asparagines at position 103 connecting TM1 and TM2 are also conserved across all *Megalgrion* species along with the other insect opsins. The DR (Aspartic acid/ Arginine) motif at position 163-164 in the third TM region involved in activation of the G-protien transduction is also conserved in all *Megalagrion* UV sequences. A triplet of amino acids that marks the consensus site for glycosylation (N–X–S/T) in *Drosophila* opsins can also be found in *Megalagrion* UV opsins at position 25. All these sites are highlighted in Figure 3.2

Tracing the amino acid substitutions important in wavelength discrimination

The interactions between the amino acid side chains and the retinal chromophore can result in shifts in absorption maxima of the opsin protein (Chang et al., 1995). Some of these specific amino acid replacements that are important in determining wavelength maxima were identified in *Megalagrion* UV opsins, which are summarized in Table 4. The replacement sites are also mapped in the phylogeny shown in Figure 3.4. Amino acid replacements that are more likely to be found within the chromophore-binding pocket are traced along the branches of the phylogenetic tree.

There are four amino acid residues described in Chang et al. (1995) that are known to interact with the chromophore to produce a diversity of absorption spectra in different opsins. Several other amino acid residues along with the four shown in Table 4 have been identified in previous studies for different model systems (Chang et al., 1995).

Table 3.4 summarizes four of the very important amino acid replacements along different lineages of the phylogeny shown in Figure 3.4. Amino acid positions 289 and 307 (numbering is based on the Bovine rhodopsin) in the TM domain VII are located about two turns of the α-helix below Lys-296, and three turns above it, respectively. The arthropods have non-polar residues at both of these positions where UV opsins have Methionine at position 289 and Alanine at position 307. All vertebrate and other arthropod opsins have a polar residue at position 113 (Chang et al., 1995). In contrast, insect UV opsins have Phenylalanine, which is a nonpolar, uncharged residue at position 113 (Chang et al., 1995). *Megalagrion* UV opsins also have a Phenylalanine at the corresponding position. Residue 124 is a polar (Ser) molecule in all the invertebrate opsins except in the UV opsins, which have a conserved nonpolar Alanine residue at this position (Chang et al.,1995). All nine species of *Megalagrion* possess all the four amino acid substitutions at the positions 113,124,289 and 307 that are conserved along the insect UV opsin lineage.

3. Test for the signatures of selection
***codeml,* PAML**

Table 4.2 shows results of Maximum Likelihood pair wise comparisons using a codon based substitution model for 9 *Megalagrion* UV opsin sequences. In the codon based substitution model, the relationship holds that $\omega = dN/dS$, where dN/dS is the ratio of non-synonymous to synonymous substitution rates. Based on the results of this analysis, the UV opsin genes of Megalagrions are evolving under purifying selection, as all the ω values are less than 1.

Table 4.3 summarizes the results of the site models which assigns variable ω ratios among sites to identify sites that are under diversifying selection (where $\omega > 1$). In total,

codeml, identified 12 potential sites which might be evolving under diversifying or positive selection.

However, the site models reveal variable selective pressures among different sites of the UV opsin protein since the ω values at these sites are either greater than 1 or very close to 1. The results show that there are a maximum of 12 amino acid sites that have a high probability of having ω values that are >1, suggesting that these sites might be under positive selection even though the whole protein is evolving under purifying selection.

4. Rate of molecular evolution

Table 8, 9 and 10 show summaries of the results of the two-cluster relative-rate tests of Takezaki et al. (1995) comparing *Megalagrion* sequences with different out groups. Table 8 shows a comparison of the nucleotide sequence data of the complete ORF of the UV opsin of different *Megalagrion* species. *Apis cerana* UV opsin sequence was used as the closest out group to compare the rate of evolution of *Megalagrion* UV opsin genes. The results show that UV opsin genes of some species are evolving at a slightly slower or slightly faster rate compared to others. However, none of the comparisons were statistically significant at the 5% level to reject the null hypothesis of equal rates of molecular evolution. *Megalagrion hawaiiense* shows a faster rate of evolution than all the other species. The rate of evolution of *M. nigrolineatum* and *M. oresitrophum* are identical and seem to evolve at a slower rate than the other *Megalagrion* species. Also *M. xanthomelas* and *M. calliphya* have equal rates of the molecular evolution.

Chapter 4

DISCUSSION

I isolated and characterized the opsins of a UV absorbing rhodopsin and a partial sequence of a long wavelength absorbing rhodopsin of the *Megalagrion* damselflies. Based on amino acid alignments with other insect opsins, the UV opsin is composed of a single exon. It seems likely, based on the phylogenetic relationships and the proven spectral sensitivities in other insects that the isolated UV opsins can absorb approximately 345-375nm wavelength light. Several attempts to isolate the long wavelength detecting opsins with different combinations of degenerate primers have remained unsuccessful so far. It is possible that the LW opsin has one or more introns making it difficult to isolate the complete LW sequence since genomic DNA from the preserved samples was used for the PCR reaction.

In a phylogenetic tree, opsins will cluster together primarily by physiological similarity i.e., similar spectral sensitivities, and secondarily by species relationships (Briscoe, 2000). Our results follow the above observation as the opsins of physiological similarity seem to cluster together across the topology of the phylogenetic tree (Figure 2.1 and 2.2). Insect opsins can be divided into three spectral classes (LW, blue, and UV) based on the results of electrophysiological experiments (Kashiyama et al., 2009). This study revealed that *Megalagrions* posses UV and LW detecting opsins, two of the main opsin types. Given that these damselflies are highly visually oriented insects that rely on visual cues for almost all of their activities, it is not surprising to find that they posses two of the main opsin types and it is likely that they would have the third opsin type also. Thus, the

results of this study suggest that *Megalagrion* damselflies have a highly developed visual system.

The topology of *Megalagrion* opsin phylogeny is consistent with the phylogeny generated by Jordan et al. (2003) using other molecular data (elongation factor α-1 and cytochrome oxidase II data) and breeding habitat usage. More detailed investigations of the relationship between changes in opsin gene sequences and breeding habitat selection would provide valuable insights in to correlations between the adaptations in wavelength discrimination and habitat use of these visually oriented insects. Therefore, this study provides the basis for future studies aimed at identifying specific amino acid substitutions leading to shifts in wavelength discrimination and subsequent shifts in habitat choice. Arthropod opsin genes have been shown to have high sequence similarity between species (Kashiyama et al., 2009, Briscoe., 2001). Deduced amino acid sequences of all *Megalagrion* opsin genes show high similarity between species and with other arthropod opsins. As described in Chang et al. (1995), the four amino acids at the positions 113, 124, 289 and 307 that are unique to insect UV opsins are conserved across all the *Megalagrion* species sequenced, which provides evidence for the evolution of all insect opsins from a common ancestor. In addition, this study has led to the identification of a number of other amino acid residues that are conserved in opsin sequences between different invertebrate and vertebrate species. However, compared to vertebrate opsins, the available information regarding invertebrate opsins is insufficient to discern the exact amino acid positions in invertebrate opsins. Therefore this analysis gives a general idea of what residues are important and how they might interact during wavelength discrimination in invertebrate opsins. These identified opsin sequences were also used in

analyses aimed at identifying critical amino acid changes that might be correlated to habitat selection by different *Megalagrion* spp.

Purifying selection acting on a specific gene indicates that the functional and structural properties of a protein are being conserved across species (Yang et al. 2000). PAML results of the analysis of *Megalagrion* UV opsin sequences based on Maximum Likelihood pair wise comparisons using a codon based substitution model, indicates purifying selection is acting on the opsin genes. This suggests that natural selection is acting on *Megalagrion* UV opsin genes to conserve the functional and structural properties of the protein, which are well adapted to the environment of a particular species.

The presence of positively selected amino acid sites suggests that the protein might be subjected to adaptive evolution and the sites that are under positive selection might have significantly higher adaptive significance compared to sites that are not under positive selection. Therefore, these sites that might be evolving under positive selection can have a greater functional significance. Adaptive evolution can be either responsible for fitness differences among proteins having polymorphic amino acids or it can make certain protein well adapted to a specific habitat (Yokoyama, 2002). As mentioned earlier, the identification of these potential amino acid replacements that are under positive selection provides the basis for further experiments involving Site-Directed Mutagenesis which would provide information on the functional importance of these positions (Yokoyama, 2002).This will confirm whether the positively selected amino acid sites are responsible for the fitness of the organism by changing the adaptive advantage of that particular

protein in a specific environment. Also further analysis such as branch-site models in codeml can be used to identify the lineages that are under variable selective pressures.

The results of the relative rate test among species, suggest that some *Megalagrion* species are evolving at a slightly faster or slower rate than others. *M. hawaiiense* and *M. oahuense* are evolving at a faster rate than all the other species while *M. nigrolineatum* and *M. oresitrophum* are evolving at a slower rate compared to all other species. According to the phylogeny with dating information presented in Jordan et al., 2003, *M. nigrolineatum* and *M. oresitrophum* speciated from a common ancestor about 3.7 MYA while the speciation of *M. hawaiiense* and *M. oahuense* dates back to 1.6 MYA. *M. hawaiiense* and *M. oahuense* are the closest relatives of each other in the *Megalagrion* opsin phylogeny, which supports the above speciation event (Figure 2.1). Also, based on Jordan et al., 2003, these two species are found to be breeding in more terrestrial habitats compared to other Megalagrions used in this study. Therefore, these results suggest that the faster rate of evolution in these two species might be the result of their shift in to more terrestrial breeding habitats. Following the same line of thought, the species that have a slower rate of evolution might be well established in their habitats compared to the species that are moving into new breeding habitats. These different rates of evolution in opsin sequences in Megalagrions indicate to another interesting aspect of the radiation of this group in the Hawaiian Islands. Based on Polhemus and Asquith, (1996), some species of *Megalagrion* have radiated into fast streams and plant leaf axils on Kauai simultaneously and about 1.5 million years later another group of species moved simultaneously to seeps and terrestrial habitats on Oahu (Polhemus and Asquith, 1996). These groups of species that show shifts in breeding habitats might be an ideal system to

test the hypothesis that more established species would have slower rates of evolution of opsin gene sequences compared to species that are still adapting to new breeding habitats. Results of such investigations would provide additional information regarding the importance of visual system adaptations in the radiation of *Megalagrion* damselflies.

REFERENCES

1. Baldwin, J.M., Schertler, G.F.X., Unger, V.M. (1997). An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J. Mol. Biol*. 272: 144–164.

2. Briscoe, A. (1998). Molecular diversity of visual pigments in the butterfly *Papilio glaucus. Naturwissenschaften.* 85: 33-5.

3. Briscoe, A. (1999). Intron splice sites of *Papilio glaucus PglRh3* corroborate insect opsin phylogeny. *Gene*. 230: 101–109.

4. Briscoe, A. (2001). Functional diversification of lepidopteran opsins following gene duplication. *Mol Biol Evol.* 18: 2270-2279.

5. Briscoe, A. (2008). Reconstructing the ancestral butterfly eye: focus on the opsins. *The Journal of Experimental Biology.* 211: 1805-1813.

6. Carulli, J. and Hartl, D.L. (1992). Variable rates of evolution among *Drosophila* opsin genes. *Genetics*. 132: 193-204.

7. Chang, B.S.W., Crandall, K.S., Carulli, J.P. and Hartl, D.L. (1995). Opsin phylogeny and evolution: a model for blue shifts in wavelength regulation. *Mol. Phy. Evol.* 4(1): 31-43.

8. Chomczynski, P. (1993). A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques*. 15: 532–534

9. Friedrich, M. (2003). Evolution of insect eye development: First insights from fruit fly, grasshopper and flour beetle. *Integr. Comp. Biol*. 43(4): 508 - 521.

10. Friedrich, M. (2006). Continuity versus split and reconstitution: Exploring the molecular developmental corollaries of insect eye primordium evolution. *Dev. Biol*. 299(2): 310-329.

11. Friedrich, M., Ying, D. and Jackowska M. (2006). Insect interordinal relationships: evidence from the visual system. *Arthropod Systematics and Phylogeny*. 64 (2): 133–148

12. Harzsch, S., Vilpoux, K., Blackburn, D.C., Platchetzki, D.,Brown, N.L., Kempler, K.E. and Battelle, B.A. (2006). Evolution of arthropod visual systems: Development of the eyes and central visual pathways in the horseshoe crab *Limulus polyphemus* Linnaeus. *Developmental Dynamics*. 235: 2641-55.

13. Hasegawa, M., Kishino, H. and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22 (2): 160–74.

14. Huelsenbeck, J. P. and Ronquist, F. (2001). MrBayes: Bayesian inference of phylogeny. *Bioinformatics*. 17: 754–755.

15. Jordan, S., Simon, C. and Polhemus, D.A. (2003). Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus *Megalagrion* (Odonata: Coenagrionidae). *Systematic Biology*: 52: 89–109.

16. Jukes, T.H. and Cantor, C.R. (1969). Evolution of protein molecules. In Munro HN, editor, *Mammalian Protein Metabolism*, 21: 132.

17. Kashiyama, K., Seki, T., Numata, H. and Goto, S.G. (2009). Molecular characterization of visual pigments in Branchiopoda and the evolution of opsins in Arthropoda. *Mol. Biol. Evol*. 26(2): 299 – 311.

18. Kimura, M. (1983). The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.

19. Koyanagi, M., Nagata, T., Katoh, K., yamashita, S. and Tokunaga, F. (2008). Molecular evolution of arthropod color vision deduced from multiple opsin genes of jumping spiders. *J. Mol. Evol*. 66(2): 130-7.

20. Kumar, S. (1996). PHYLTEST: A program for testing phylogenetic hypothesis Version 2.0. Institute of Molecular and Evolutionary Genetics and Department of Biology, The Pennsylvania State University, University Park, Pennsylvania.

21. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan., P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21): 2947-2948.

22. Maddison, W. and Maddison, D. (2003). MacClade, v. 4.06. Sinauer Associates.

23. Polhemus, D. and Asquith, A. (1996) . Hawaiian Damselflies, A field identification guide. Bishop Museum press, Honolulu, Hawaii.

24. Polhemus, D. (1997). Phylogenetic analysis of the Hawaiian damselfly genus *Megalagrion* (Odonata: Coenagrionidae): implications for biogeography, ecology and conservation biology. *Pacific Science.* 51(4): 395–412.

25. Salcedo, E., Farrell, D.M., Zheng, L., Phistry, M., Bagg, E. and Britt, S.G. (2009). The green absorbing Drosophila Rh6 visual pigment contains a blue shifting amino acid substitution that is conserved in vertebrates. *J. Biol. Chem*. 284(9): 5717-22.

26. Spady, T., Seehausen, O., Loew, E., Jordan, R., Kocher, T. and Carleton, K. (2005). Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Mol. Biol. Evol*. 22(6): 1412-22.

27. Spaethe, J. and Briscoe, A. (2004). Early duplication and functional diversification of the opsin gene family in insects. *Mol. Biol. Evol*. 21: 1583-1594.

28. Stavenga, D. (2006). Invertebrate photoreceptor optics. Invertebrate Vision. Cambridge University Press, 1-42.

29. Swofford, D. L. (2002). PAUP*. Phylogenetic Analysis Using Parsimon.Version 4. Sinauer Associates, Sunderland, Massachusetts.

30. Tajima, F. (1993). Simple methods for testing molecular clock hypothesis. *Genetics* 135: 599-607.

31. Takahashi, Y., Yokoyama, S. (2005). Genetic basis of spectral tuning in the violet-sensitive visual pigment of African clawed frog, *Xenopus laevis. Genetics Society of America*. 171: 1153-1160

32. Takizaki, N., Razhetsky, A. and Nei, M. (1995). Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol*. 12: 823–833.

33. Taylor, S. D., Cruz, K.D., Porter, M.L. and M.F. (2005) Characterization of the long wavelength opsin from Mecoptera and Siphonaptera: Does a flea see? *Mol. Biol. Evol.* 22: 1165-1174.

34. Terakita, A. (2005). The opsins. *Genome Biol*. 6: 213.

35. Thompson, E. (1995). Color Vision: A study in cognitive science and the philosophy of perception. Routledge. New York. Publication.

36. Towner, P. and Giirtnerb, W.(1994) The primary structure of mantid opsin *Gene*, 143: 227-231.

37. Townson, S.M., Chang, B. S.W., Salcedo, E., Chadwell, L., Pierce, N. E. and Britt, S. G. (1998). Honeybee blue- and ultraviolet-sensitive opsins: Cloning, heterologous expression in *Drosophila*, and physiological characterization. *J. Neurosci.* 18: 2412-2422.

38. Yang, Z., Swanson, W.J., Vacquier, V.D. (2000). Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol. Biol. Evol.* 17: 1446–1455.

39. Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24: 1586-1591.

40. Yokoyama, S. (2002). Evaluating adaptive evolution. *Nature genetics* .30. Nature Publishing Group. http://genetics.nature.com

FIGURE LEGENDS

Figure 1: Schematic diagram of a rhodopsin molecule showing the opsin moiety and the chromophore (11-cis-retinal) embedded within the transmembrane domains. http://www2.fz-juelich.de/isb/isb-1/Rhodopsin

Figure 2.1: Bayesian tree of *Megalagrion* damselfly species relationships reconstructed using the mixed-model for amino acid sequences in MrBayes 3.1. Branch labels indicate the Bayesian posterior probabilities based on a 1 million step MCMC analysis. The cluster highlighted in blue includes UV opsins of *Megalagrion* and *Apis* sp. The cluster highlighted in red includes the LW opsin sequences of *Megalagrion* spp.

Figure 2.2: Bayesian tree of *Megalagrion* damselfly species reconstructed using the HKY model for nucleotide sequences in MrBayes 3.1. Branch labels indicate the Bayesian posterior probabilities based on a 1 million step MCMC analysis. Blue and red highlighted clusters correspond to the *Megalagrion* UV and LW clades respectively.

Figure 3.1: CLUSTAL W 2.0.12 multiple sequence alignment of the deduced amino acids of the UV opsin genes of nine species of Megalagrions and other closely related insect UV opsins. Shaded blocks indicate the transmembrane domains (I–VII). A known functionally important five amino acid residues conserved across the UV opsin clade are highlighted in yellow.

Figure 3.2: Nucleotide and deduced amino acid sequence of *Megalagrion* UV opsin. Nucleotides are numbered in the 5'- 3'direction.The cDNA with UTRs is 1342 bp long and codes for a protein of 387 amino acids. Numbers on the right-hand side are for amino acids (in bold) and for nucleotides (regular). The grey boxes designate seven potential transmembrane (TM) domains. The DRY motif is a site of G-protein binding/activation (Baldwin et al., 1997). Potential sequences for G-protein binding sites, DRY and QAKKMNV are indicated by open boxes. The seven amino acid motifs that are important for a functional opsin protein that are conserved across all known insect opsins are highlighted in yellow and by open boxes.

Figure 3.3: A schematic diagram of the *Megalagrion* UV opsin protein designed using Adobe Illustrator. This diagram shows the seven transmembrane domains along with the cytoplasmic and extracellular domains. The non-synonymous substitutions are highlighted in different colors. Position corresponding to the Schiff base counter-ion in vertebrate opsins is marked light blue. Lysine (K) is the residue for chromophore attachment. Cysteines which form the disulfide bonds are also shown in the diagram. Amino acid sites are numbered based on the Bovine rhodopsin.

Figure 3.4: A simplified phylogeny showing the four conserved amino acids mapped on to phylogeny constructed by Chang et al., 1995. This phylogeny is mainly focused on the putative amino acid replacements that contribute to UV shifts in opsins. Convergent

nonconservative amino acid replacements in the positions associated with the chromophore in the opsin molecules are highlighted in this phylogeny. These amino acids can directly affect the absorption spectra of the opsin molecule (Chang et al., 1995). Ancestral states of these amino acid substitutions are given at the nodes of the tree. The four amino acid residues at the positions 124, 113, 289 and 307 are conserved across all *Megalagrion* UV opsin sequences.

Figure 5.1: An illustration showing the comparison of data under the two-cluster relative-rate test as implemented in PHYLTEST software (Kumar 1996).
Assume that L_a and L_b are the mean branch length or the averages of observed numbers of substitutions per site (branch lengths) from the common ancestor of clusters A and B. If $L_a = L_b$, the null hypothesis is accepted under the constancy of molecular clock, since $L_a - L_b = 0$ indicating no difference in the number of substitutions in between the two lineages. In that case species A and B are evolving at a same rate.

LIST OF TABLES

Table 2.1: UV and LW gene specific and degenerate primers which were used in PCR.

Table 3.1: Conserved amino acid residues that are important for specific wavelength detection.
This table summarizes four of the very important amino acid replacements along different lineages of the phylogeny, previously described by Chang et al. (1995). Amino acid positions are numbered according to the numbering of bovine rhodopsin.

Table 4.1: Parameters used in Site Models as implemented in *codeml*, PAML.
Model M0 is the null model with no variation among sites. M1a is the nearly neutral model with two categories of sites. M2 is known as the "selection" model with three categories of sites while M3, the ''discrete'' model also has three categories of sites, where the dN/dS ratio is free to vary for each site. M7 model uses 10 categories of sites with 10 dN/dS ratios. M8 uses 10 categories of sites and an additional category with omega that is free to vary (Yang 2007). The results of the site models are summarized in Table 4.3.

Table 4.2: Maximum Likelihood pair wise comparison of dN/dS ratios of the UV opsin sequences of nine different *Megalagrion* species. Analysis was performed using a codon based substitution model of codeml, PAML (Nei & Gojobori 1986).

Table 4.3: Parameters and Log-Likelihood Values in the Site Models of Variable ω Ratios among sites.
Table 4.3 is a summary of the application of variable ω ratios among sites to find the presence of sites under diversifying selection (where ω >1). A site refers to an amino acid or codon rather than a nucleotide throughout the analysis. This table lists parameter estimates and log-likelihood values under models of variable ω ratios among sites. PAML applies six different models in the CODEML analysis using the variable **Nssites**. The models are discussed in the method section. No sites were identified as positively selected under M0, M1a and M7 models. M2a identified 5 sites which are under positive selection. M3 (discrete) involves four more parameters than M0 (one ratio). All the models (M2, M3 and M8) that allow for the presence of positively selected sites do identify sites under positive selection. All together CODEML identified 12 potential sites which evolve under diversifying or positive selection.

Table 4.4: Likelihood Ratio Statistics comparing different site models. This table lists parameter estimates and log-likelihood values (LRT) under models of variable ω ratios among sites. The results shown in table 4.3 were used to calculate the LRT values as shown in Table 4.4. The LRT values were then compared with the chi-squared values and if the LRT statistic is greater than the critical value of chi-squared, the results suggest extreme variation in selective pressure among amino acid sites. For instance, in the comparison of M0 and M3 models, the LRT statistic of 57.52 is greater than the chi-squared value of 13.28 when df=4. These results suggest a variation in selective pressure

among amino acid sites. LRTs greater than the chi-squared values for all comparisons indicate that there is a variation of selective pressure between amino acids in the *Megalagrion* UV opsins and might be evolving under positive selection.

Table 5.1: Cluster assignments in each analysis of relative rate test. This table shows how the three clusters as shown in figure 5.1 were specified based on the comparison. Cluster A and B are the in-groups to be compared when an out group C is provided. PHYLTEST sets the equal rate of evolution or the consistency of the rate of evolution as the null hypothesis. PHYLTEST then simply tests for the acceptance or the rejection of the null hypothesis under the constancy of molecular clock. Constancy of the molecular clock in between two lineages (A and B) is tested when an out group lineage (C) is given.

Table 5.2: Comparison of the rates of evolution among different *Megalagrion* species. The symbol "S" indicates that the species in column Y exhibited a slower rate of evolution than that of the species in row X
"F" indicates that the species in column Y exhibited a faster rate of evolution.
"*" next to the symbol indicates that the difference in evolutionary rates for that comparison was statistically significant at the 5% level by a two-tailed normal deviate test conducted in PHYLTEST.
"I" indicates that the rates were identical for both groups. UV opsin sequence of *Apis cerana* was used as the out group.

Table 5.3: Comparison of the rates of evolution in between UV and LW sequences of different *Megalagrion* species. The symbol "S" indicates that the species in column Y exhibited a slower rate of evolution than that of the species in row X
"F" indicates that the species in column Y exhibited a faster rate of evolution.
"*" next to the symbol indicates that the difference in evolutionary rates for that comparison was determined to be statistically significant at the 5% level by a two-tailed normal deviate test conducted in PHYLTEST.
"I" indicates that rates were identical for both groups. UV opsin sequence of *Apis cerana* was used as the out group.

Table 5.4: Comparison of the rates of evolution among different domains of the UV opsin protein of different *Megalagrion* species. The rate of evolution of TM (trans-membrane) and NTM (non transmembrane) domains of the opsin proteins were compared with the out group ORF (complete open reading frame).
"F" indicates that the species in column Y has a faster rate of evolution.
"*" next to the symbol indicates that the difference in evolutionary rates for that comparison was determined to be statistically significant at the 5% level by a two-tailed normal deviate test conducted in PHYLTEST.
"I" indicates that rates were identical for both groups. NA indicates that a sequence from the indicated ingroup was not available for comparison.
UV opsin sequence of *Apis cerana* was used as the out group.

FIGURES

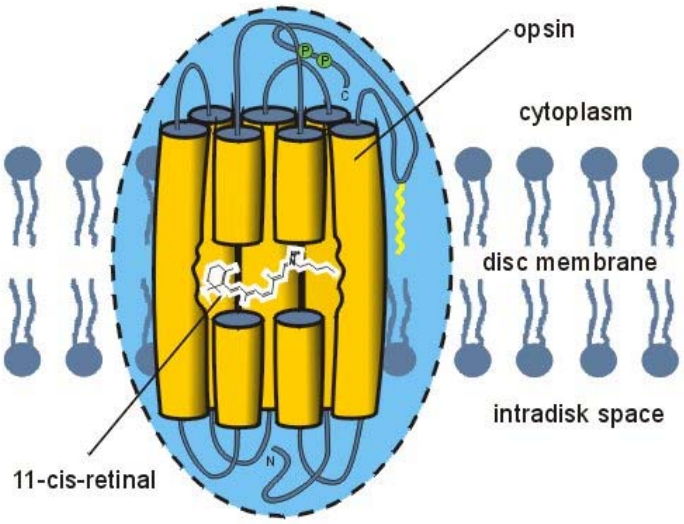


Figure 1

Figure 2.1

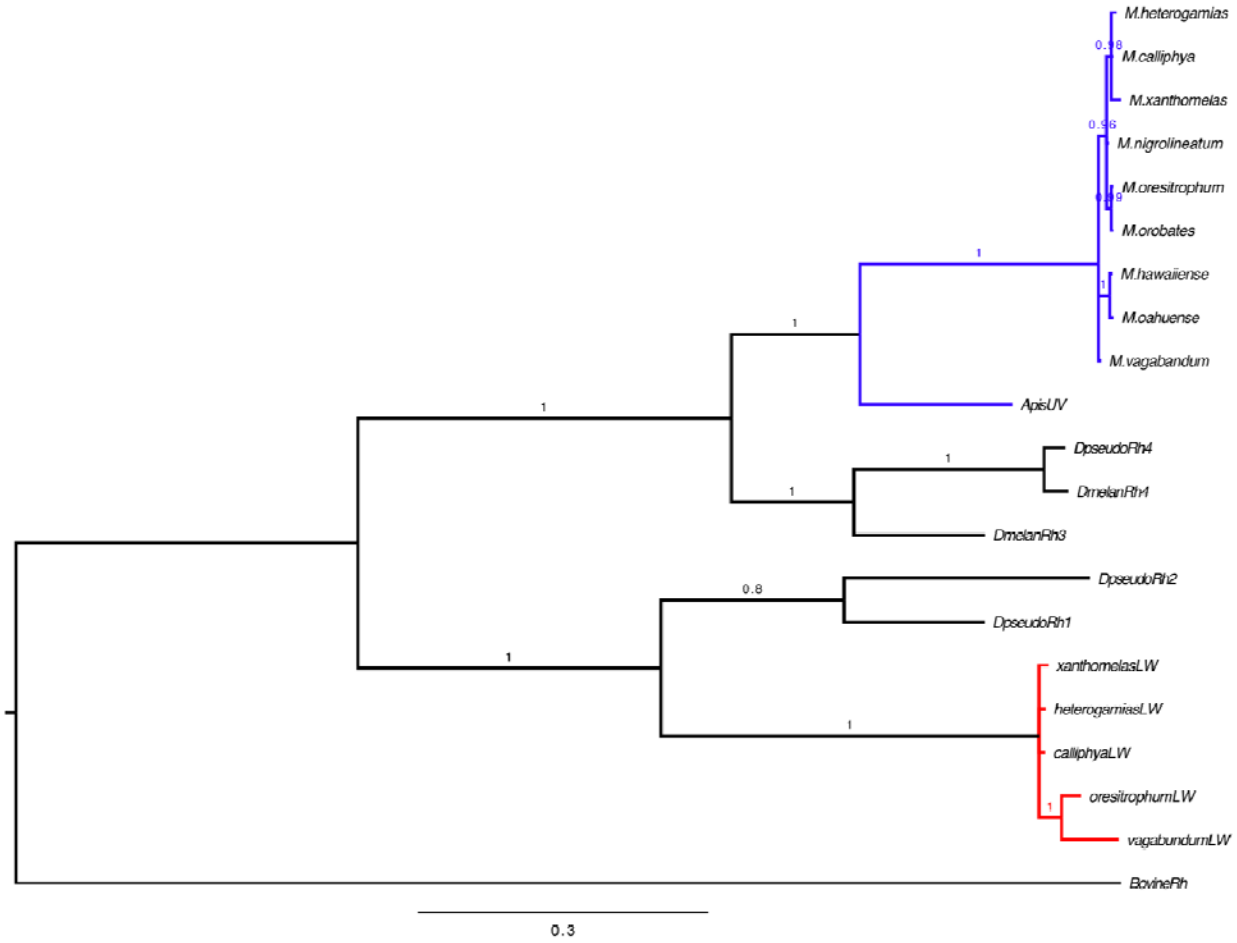
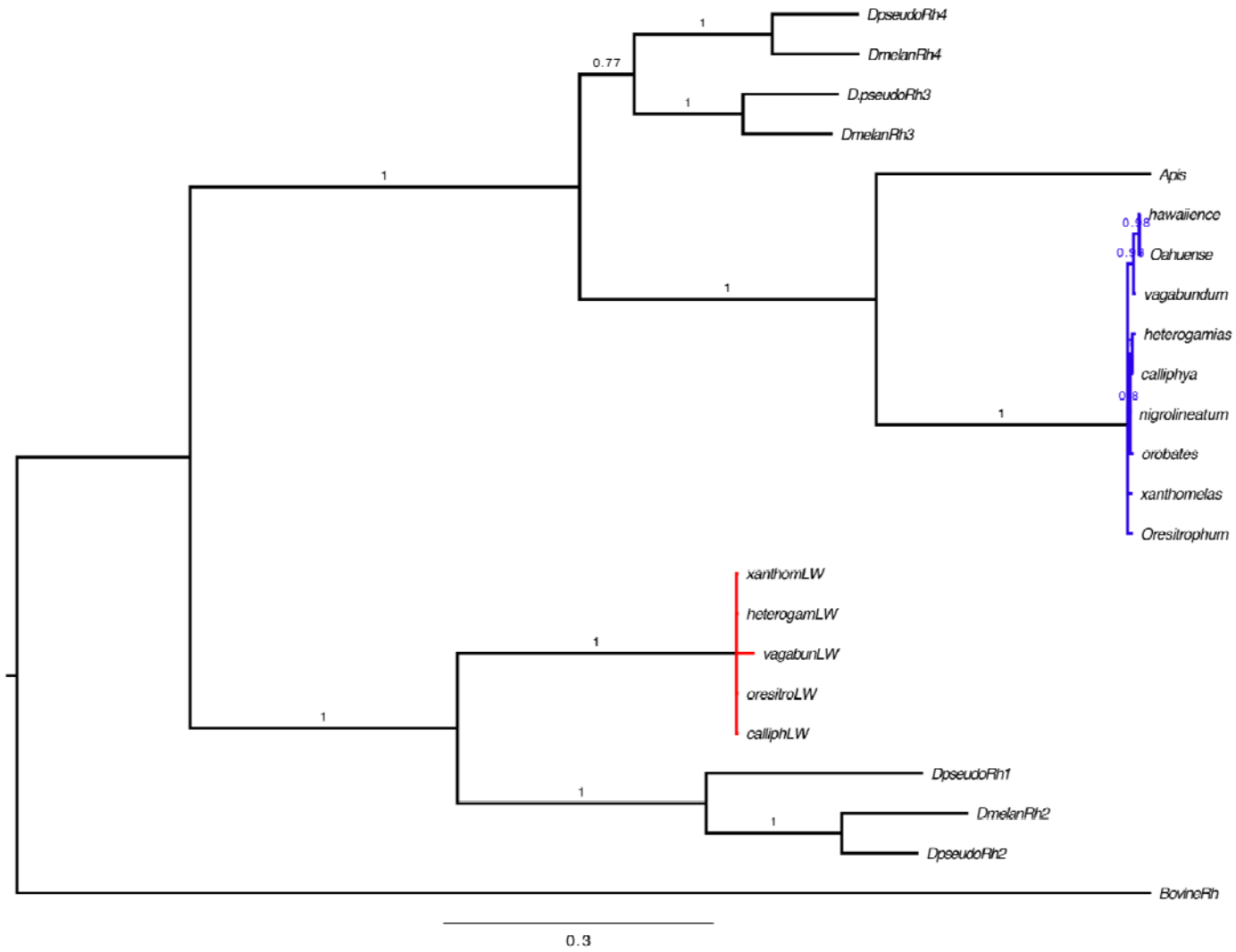
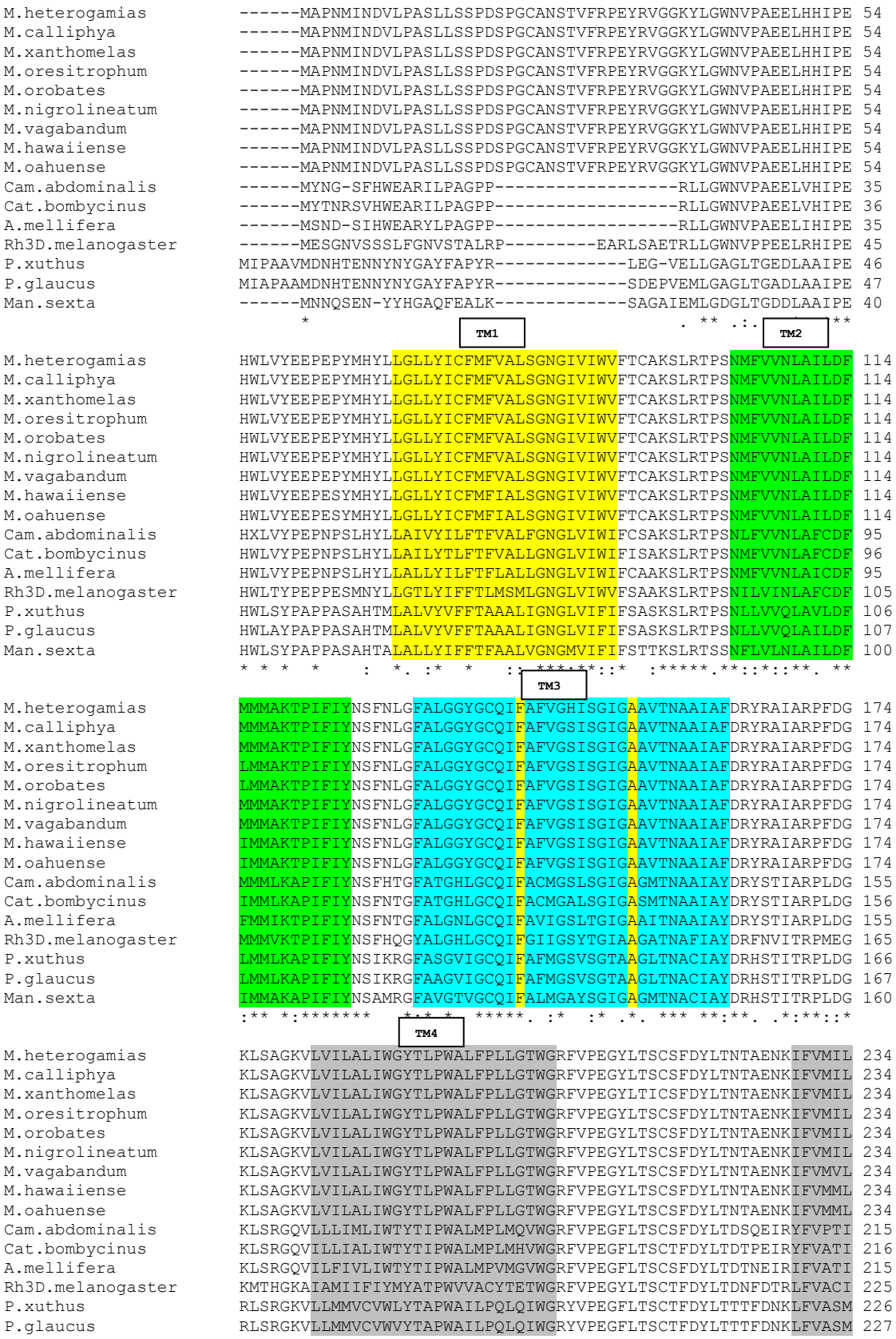

M.heterogamias
M.calliphya
M.xanthomelas
M.nigrolineatum
M.oresitrophum
M.orobates
M.hawaiiense
M.oahuense
M.vagabandum
ApisUV
DpseudoRh4
DmelanRh1
DmelanRh3
DpseudoRh2
DpseudoRh1
xanthomelasLW
heterogamiasLW
calliphyaLW
oresitrophumLW
vagabundumLW
BovineRh
0.3

Figure 2.2


DpseudoRh4
DmelanRh4
D.pseudoRh3
DmelanRh3
Apis
hawaiiense
Oahuense
vagabundum
heterogamias
calliphya
nigrolineatum
orobates
xanthomelas
Oresitrophum
xanthomLW
heterogamLW
vagabunLW
oresitroLW
calliphLW
DpseudoRh1
DmelanRh2
DpseudoRh2
BovineRh
0.77
1
0.3


M.heterogamias
M.calliphya
M.xanthomelas
M.oresitrophum
M.orobates
M.nigrolineatum
M.vagabandum
M.hawaiiense
M.oahuense
Cam.abdominalis
Cat.bombycinus
A.mellifera
Rh3D.melanogaster
P.xuthus
P.glaucus
Man.sexta
TM1
TM2
TM3
TM4

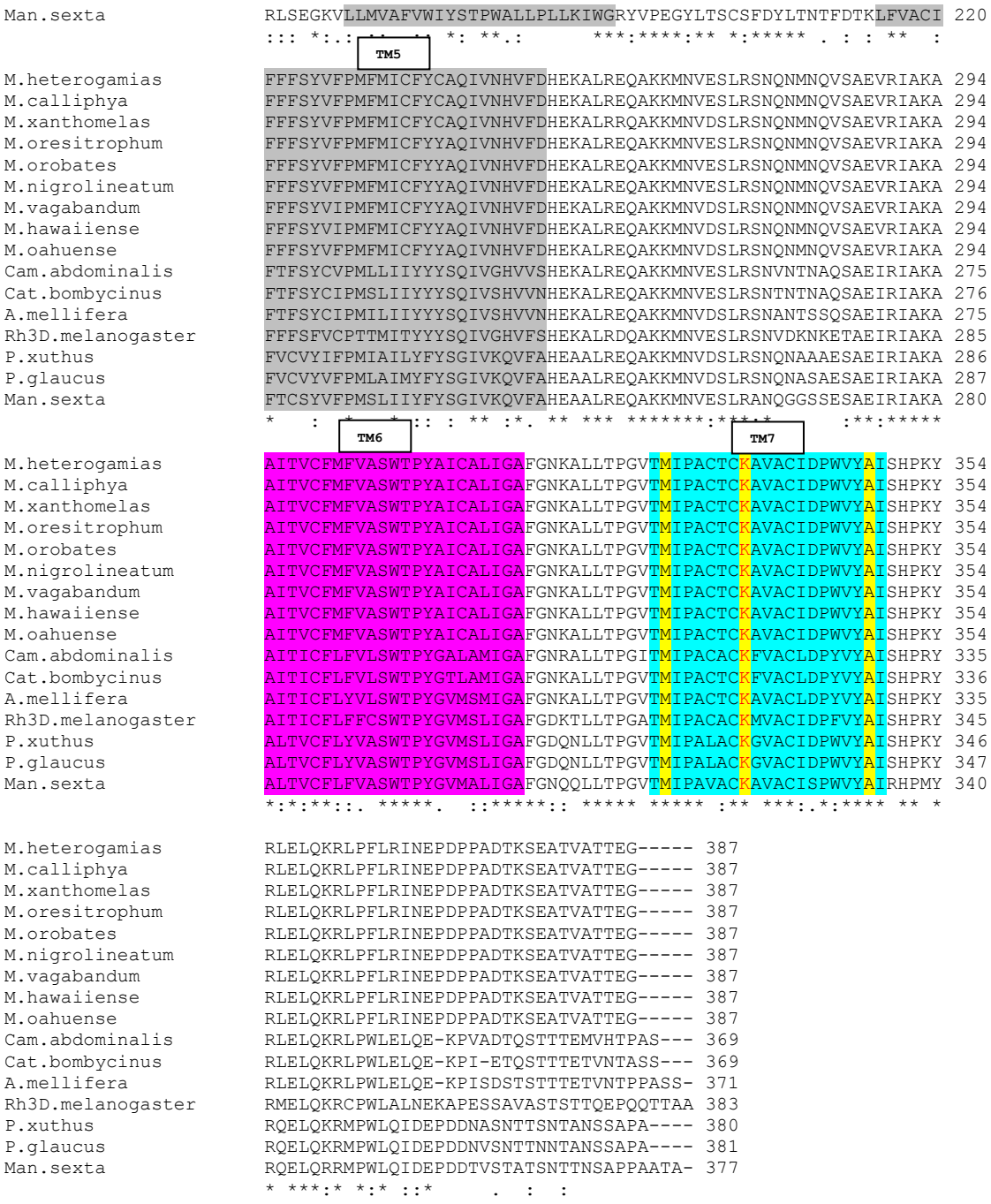


Figure 3.1

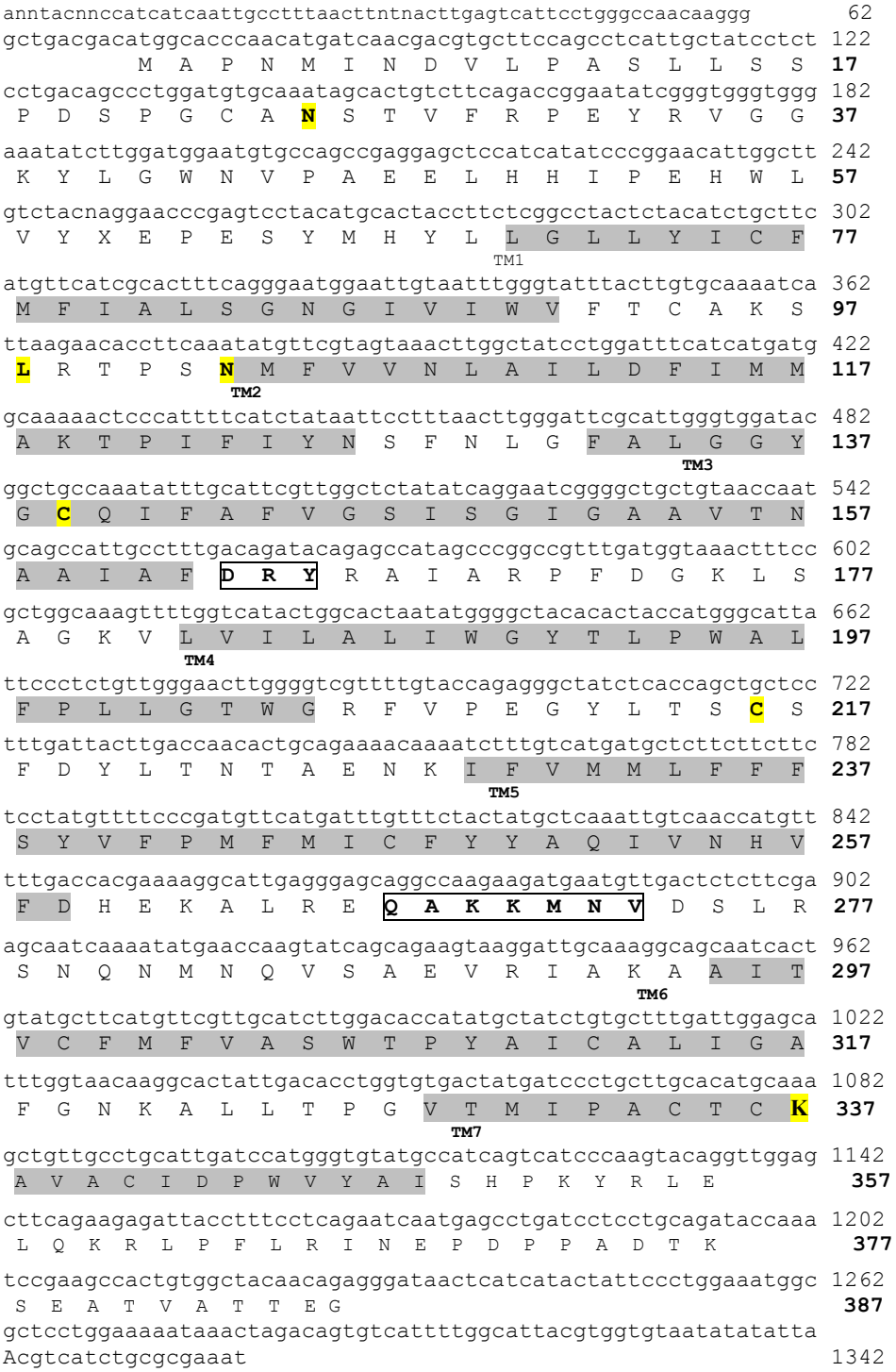

anntacnnccatcatcaattgcctttaacttntnacttgagtcattcctgggccaacaaggg 62
gctgacgacatggcacccaacatgatcaacgacgtgcttccagcctcattgctatcctct 122
M A P N M I N D V L P A S L L S S 17
cctgacagccctggatgtgcaaatagcactgtcttcagaccggaatatcgggtgggtggg 182
P D S P G C A N S T V F R P E Y R V G G 37
aaatatcttggatggaatgtgccagccgaggagctccatcatatcccggaacattggctt 242
K Y L G W N V P A E E L H H I P E H W L 57
gtctacnaggaacccgagtcctacatgcactaccttctcggcctactctacatctgcttc 302
V Y X E P E S Y M H Y L L G L L Y I C F 77
TM1
atgttcatcgcactttcagggaatggaattgtaatttgggtatttacttgtgcaaaatca 362
M F I A L S G N G I V I W V F T C A K S 97
ttaagaacaccttcaaatatgttcgtagtaaacttggctatcctggatttcatcatgatg 422
L R T P S N M F V V N L A I L D F I M M 117
TM2
gcaaaaactcccattttcatctataattcctttaacttgggattcgcattgggtggatac 482
A K T P I F I Y N S F N L G F A L G G Y 137
TM3
ggctgccaaatatttgcattcgttggctctatatcaggaatcggggctgctgtaaccaat 542
G C Q I F A F V G S I S G I G A A V T N 157
gcagccattgcctttgacagatacagagccatagcccggccgtttgatggtaaactttcc 602
A A I A F D R Y R A I A R P F D G K L S 177
gctggcaaagttttggtcatactggcactaatatggggctacacactaccatgggcatta 662
A G K V L V I L A L I W G Y T L P W A L 197
TM4
ttccctctgttgggaacttggggtcgttttgtaccagagggctatctcaccagctgctcc 722
F P L L G T W G R F V P E G Y L T S C S 217
tttgattacttgaccaacactgcagaaaacaaaatctttgtcatgatgctcttcttcttc 782
F D Y L T N T A E N K I F V M M L F F F 237
TM5
tcctatgttttcccgatgttcatgatttgtttctactatgctcaaattgtcaaccatgtt 842
S Y V F P M F M I C F Y Y A Q I V N H V 257
tttgaccacgaaaaggcattgagggagcaggccaagaagatgaatgttgactctcttcga 902
F D H E K A L R E Q A K K M N V D S L R 277
agcaatcaaaatatgaaccaagtatcagcagaagtaaggattgcaaaggcagcaatcact 962
S N Q N M N Q V S A E V R I A K A A I T 297
TM6
gtatgcttcatgttcgttgcatcttggacaccatatgctatctgtgctttgattggagca 1022
V C F M F V A S W T P Y A I C A L I G A 317
tttggtaacaaggcactattgacacctggtgtgactatgatccctgcttgcacatgcaaa 1082
F G N K A L L T P G V T M I P A C T C K 337
TM7
gctgttgcctgcattgatccatgggtgtatgccatcagtcatcccaagtacaggttggag 1142
A V A C I D P W V Y A I S H P K Y R L E 357
cttcagaagagattacctttcctcagaatcaatgagcctgatcctcctgcagataccaaa 1202
L Q K R L P F L R I N E P D P P A D T K 377
tccgaagccactgtggctacaacagagggataactcatcatactattccctggaaatggc 1262
S E A T V A T T E G 387
gctcctggaaaaataaactagacagtgtcattttggcattacgtggtgtaatatatatta
Acgtcatctgcgcgaaat 1342

Figure 3.2

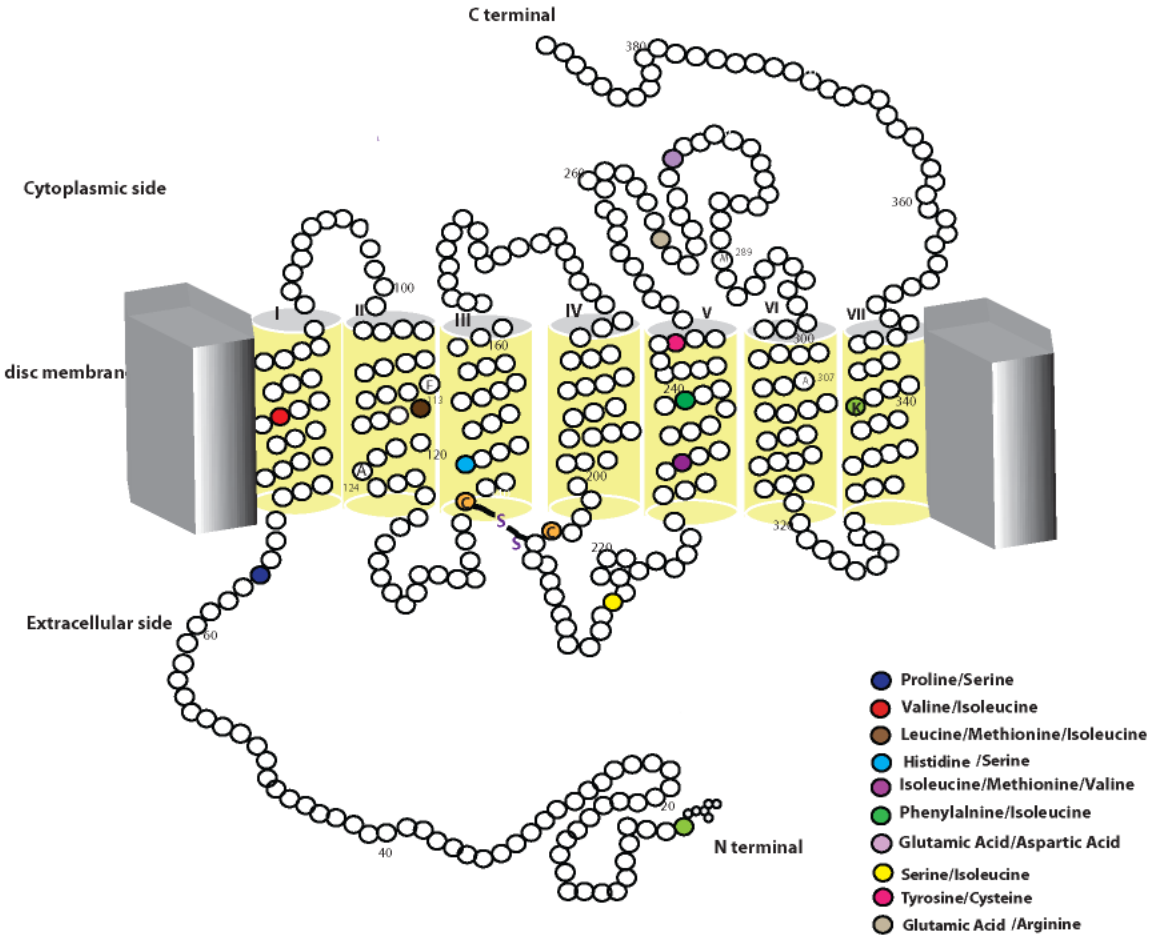

C terminal
Cytoplasmic side
disc membran
Extracellular side
N terminal
I
II
III
IV
V
VI
VII
Proline/Serine
Valine/Isoleucine
Leucine/Methionine/Isoleucine
Histidine /Serine
Isoleucine/Methionine/Valine
Phenylalnine/Isoleucine
Glutamic Acid/Aspartic Acid
Serine/Isoleucine
Tyrosine/Cysteine
Glutamic Acid /Arginine

Figure 3.3

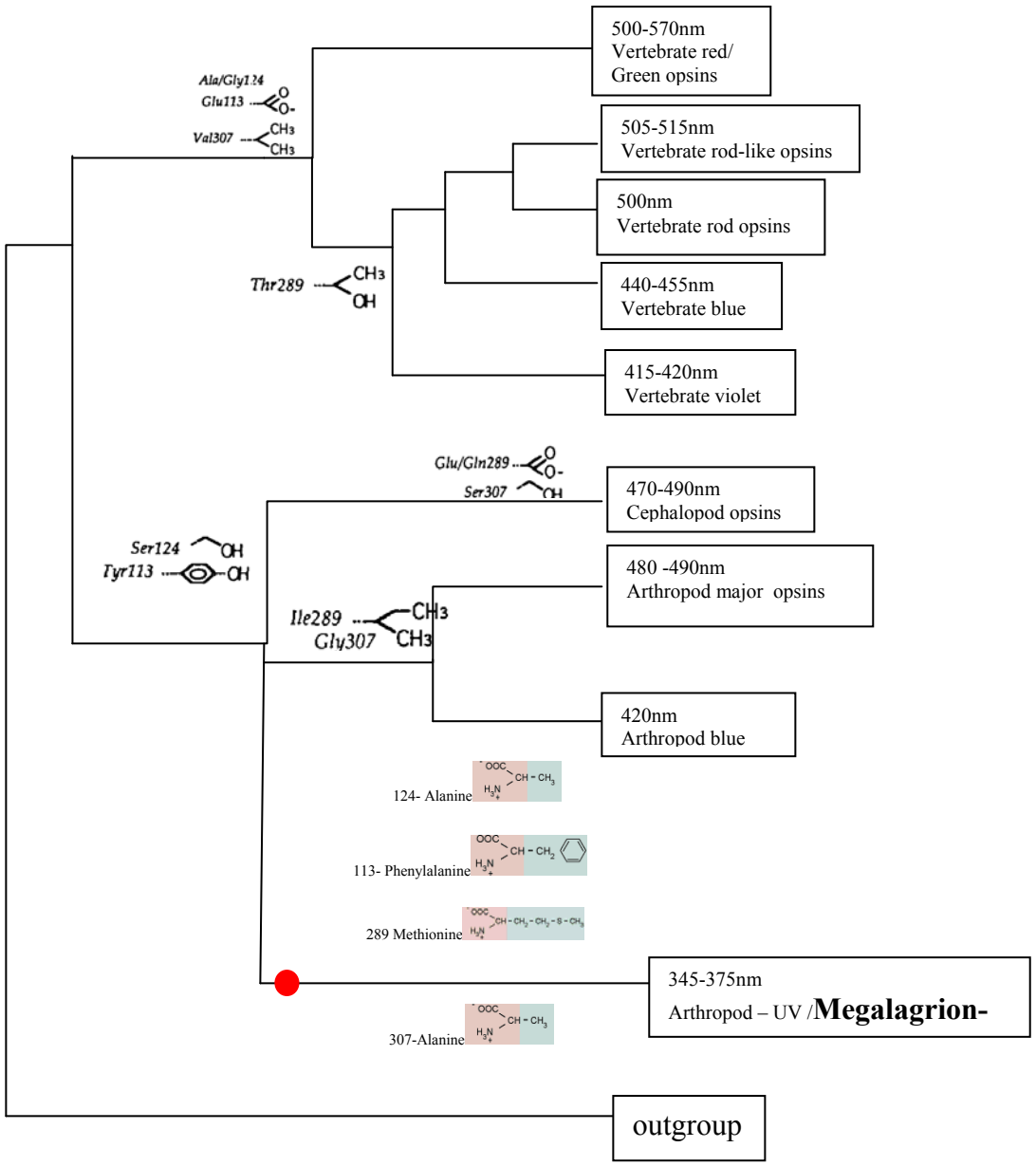

500-570nm
Vertebrate red/
Green opsins
Ala/Gly124
Glu113
Val307
505-515nm
Vertebrate rod-like opsins
500nm
Vertebrate rod opsins
Thr289
440-455nm
Vertebrate blue
415-420nm
Vertebrate violet
Glu/Gln289
Ser307
470-490nm
Cephalopod opsins
Ser124
Tyr113
480 -490nm
Arthropod major opsins
Ile289
Gly307
420nm
Arthropod blue
124- Alanine
113- Phenylalanine
289 Methionine
345-375nm
Arthropod – UV /Megalagrion-
307-Alanine
outgroup

Figure 3.4

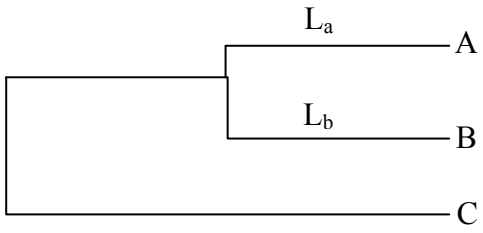


Figure 5.1

TABLES

Table 2.1

Primer	Type	Opsin class	Sequence (5'-3')
UV1F	Gene specific	UV	GAGGAAAGCTAGGAAAGAGC
UV1R	Gene specific	UV	CACCCAACATGATCAACGAC
UVM1R	Gene specific	UV	ATGGCACCCAACATGATCAACG
UVM1R	Gene specific	UV	TCTTGGCCTGCTCCCTCAAT
UVM2F	Gene specific	UV	GCCTGCTCCCTCAAT
UVM2R	Gene specific	UV	TGTTCATCGCACTTTCA
UVM3F	Gene specific	UV	GCTCCTTTGATTACTTGAC
UVM3R	Gene specific	UV	GGGAATAGTATGATGAGTT
LW1F	Gene specific	LW	CTACAACTGGAATAATGGCAACC
LW1R	Gene specific	LW	ATGAATGGYCCGAGGGA
LW4F	Degenerate	LW	GARCARGCNAARAARATGAAYGT
LW4R	Degenerate	LW	ARRTANGGNGTCCAHGCCA

Table 3.1

	Vertebrate					Cephalopod	Arthropod			
Position according to Bovine Rh	red/green (LW) 500-570nm	rod-like 505-515nm	Rod 500nm	blue 440-455nm	violet 415-420nm	opsins 470-490nm	major 480-490nm	Blue 420nm	UV 345-375nm	*Megalagrion* UV
113	Glu	Glu	Glu	Glu	Glu	Tyr	Tyr	Tyr	Phe	Phe
124	Ala/Gly	Ala/Gly	Ala/Gly	Ala/Ser	Thr	Ser	Ser	Ser	Ala	Ala
289	Ala	Ala	Ala	Ser/Thr	Thr	Glu/Gln	Ile	Ile	Met	Met
307	Val	Val	Ile	Val	Cys	Ser	Gly	Gly	Ala	Ala

Table 4.1

Model	NSsites	*P*	Parameters
M0 (one ratio)	0	1	ω = 0.376
M1a (neutral)	1	2	*p0 (p1 = 1 – p0),* ω0< 1, ω1= 1 p0=0.84595, ω = 0
M2a (selection)	2	4	*p0, p1 (p2 = 1 – p0 – p1),* ω0< 1, ω1= 1, ω2> 1 *p*0 = 0.95945, ω = 0 *p*1 = 0.00000, ω = 1
M3 (discrete)	3	5	*p0, p1 (p2 = 1 – p0 – p1)* ω0, ω1, ω2
M7 (beta)	7	2	*p, q* p=0.005 q=0.047
M8 (beta&ω)	8	4	*p0*=0.959, p=0.005, q=3.119, p1=0.040, w= 10.851 ***(p1 = 1 – p0),*** ***p, q, ωs> 1***

Table 4.2

Species	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. *oahuense*									
2. *hawaiience*	0.3051								
3. *nigrolineatum*	0.1356	0.1794							
4. *Oresitrophum*	0.1241	0.1629	0.4556						
5. *orobates*	0.1243	0.1631	0.1826	0.3645					
6. *vagabundum*	0.2523	0.2425	0.1404	0.2659	0.1672				
7. *calliphya*	0.1488	0.1899	0.3045	0.4244	0.2170	0.1590			
8. *xanthomelas*	0.2180	0.2705	0.7639	0.8531	0.5674	0.2621	0.4060		
9. *heterogamias*	0.1692	0.2055	0.3030	0.3878	0.2473	0.1849	0.3035	0.3635	

Table 4.3

Model	NSsites	*p*	Parameters	lnL	*d*N/*d*S	Positively Selected Sites
M0 (one ratio)	0	1	ω = 0.376	-1884.74	0.3759	None
M1a (neutral)	1	2	*p0 (p1 = 1 – p0),* ω0< 1, ω1= 1 p0=0.84595, ω = 0	-1870.27	0.1540	Not allowed
M2a (selection)	2	4	*p0, p1 (p2 = 1 – p0 – p1),* ω0< 1, ω1= 1, ω2> 1 *p*0 = 0.95945, ω = 0 *p*1 = 0.00000, ω = 1	-1855.98	0.440	115**, 147*, 233*, 250*, 266**
M3 (discrete)	3	5	*p0, p1 (p2 = 1 – p0 – p1)* ω0, ω1, ω2	-1855.98	0.440	58, 64, 80,115,147,178,215,233 ,241,250,266, 274 (all sites with P>99%)
M7 (beta)	7	2	*p, q* p=0.005 q=0.047	-1871.07	0.100	Not allowed
M8 (beta&ω)	8	4	*p0*=0.959, p=0.005, q=3.119, p1=0.040, w= 10.851 ***(p1 = 1 – p0), p, q, ωs> 1***	-1855.99	0.440	80⋆, 115**, 147⋆, 233⋆, 241⋆, 250⋆, 266**, 274⋆

Table 4.4

Comparison	$2\Delta \ell$	df	x21%
M0 (one ratio) vs. M3 (discrete)	57.52	4	13.28
M1a (neutral) vs. M2a (selection)	28.58	2	9.21
M7 (beta) vs. M8 (beta & v)	30.16	2	9.21

Table 5.1

Compare the rate of evolution	Cluster A	Custer B	Cluster C	Spp
among different domains of UV opsin	TM domains	NTM domains	Complete open reading frame of *Megalagrion* and *Apis cerana*	All the 9 *Megalagrion* spp
in between UV and LW	Segment of UV opsin	Segment of LW opsin	Segment of *Apis* UV opsin	All available *Megalagrion* sequences
Among different species	ORF of UV opsin	ORF of UV opsin	ORF of *Apis* UV opsin	All the 9 *Megalagrion* spp

Table 5.2

X / Y	*M.hawaiiense*	*M.nigrolineatum*	*M.oresitrophum*	*M.orobates*	*M.oahuense*	*M.heterogamias*	*M.calliphya*	*M.xanthomelas*	*M.vagabandum*
M.hawaiiense	-	F	F	F	F	F	F	F	F
M.nigrolineatum	S	-	I	S	S	S	S	S	S
M.oresitrophum	S	I	-	S	S	S	S	S	S
M.orobates	S	F	F	-	S	S	S	S	F
M.oahuense	S	F	F	F	-	S	F	F	F
M.heterogamias	S	F	F	F	F	-	F	F	F
M.calliphya	S	F	F	F	S	S	-	I	F
M.xanthomelas	S	F	F	F	S	S	I	-	F
M.vagabandum	S	F	F	S	S	S	S	S	-

Table 5.3

UV / LW	*M.hawaiiense*	*M.nigrolineatum*	*M.oresitrophum*	*M.orobates*	*M.oahuense*	*M.heterogamias*	*M.calliphya*	*M.xanthomelas*	*M.vagabandum*
M.oresitrophum	F*	F*	F*	F*	F*	F*	F*	F*	F*
M.heterogamias	F*	F*	F*	F*	F*	F*	F*	F*	F*
M.calliphya	F*	F*	F*	F*	F*	F*	F*	F*	F*
M.xanthomelas	F*	F*	F*	F*	F*	F*	F*	F*	F*
M.vagabandum	F*	F*	F*	F*	F*	F*	F*	F*	F*

Table 5.4

	UV opsin *Megalagrion* NTM domains (UV opsin ORF as the out group)	UV opsin *Apis cerana* NTM domains (UV opsin *Apis cerana* ORF as the out group)	UV opsin *Megalagrion* NTM domains (UV opsin *Apis cerana* ORF as the out group)
UV opsin *Megalagrion* TM domains	F*	NA	F
UV opsin *Apis cerana* TM domains	NA	F*	NA