ABSTRACT

Hill, Hilary Nan. Phylogenetic Utility of Two New Nuclear Genes, Opsin and CAD, within the stiletto flies (Diptera: Therevidae) (Under the direction of Brian M. Wiegmann).

A need for multiple phylogenetic markers to reconstruct evolutionary relationships is increasingly apparent within both the Therevidae and insect systematics as a whole, especially markers that can accommodate the immense species diversity that arose during the Mesozoic (~65-250mya). Phylogenetic trees are often best reconstructed using datasets from distinct sources and from character sets that have been under different selective pressures. Many recent insect systematics studies use various combinations of markers from multiple genomes, morphology, and biogeography. Nuclear genes, particularly protein-encoding genes that are highly conserved and low copy-number, are increasingly attractive as phylogenetic markers, and there is a recent concerted effort to find and increase access to new nuclear genes. Chapter one of this study investigates the phylogenetic utility of opsin, a new multiple copy nuclear protein-encoding gene in the dipteran family Therevidae. The present analyses of nucleotide sequence data of opsin reconstructed a therevid phylogeny that is partially concordant with previous classifications which demonstrates that opsin may have some limited utility as a phylogenetic marker within the Therevidae and closely related Diptera, but also shows that opsin may be better used in combination with other molecular markers or morphological data sets. Phylogenetic analyses of opsin at multiple levels within insects supports these results, but also reveals the significant obstacles in technical manipulation of the gene and in the interpretaion of ortholog/paralog relationships. In the second chapter two new nuclear, protein-encoding genes, opsin and CAD, in combination with EF-1α and 28S rDNA, are applied to reconstruct evolutionary relationships among the major lineages of Therevidae for which previous molecular evidence has been insufficient. The analyses of 28S rDNA, EF-1α, opsin, and CAD nucleotide sequences provide phylogenetic resolution that is highly concordant with the current therevid classification. The analyses also reveal that the phycine tribe Xestomyzini is monophyletic and

could represent a major therevid clade. The new molecular dataset generated from this study can be used to build a larger dataset and can also be used in conjunction with a morphological dataset that is currently being generated within the therevid PEET project.

PHYLOGENETIC UTILITY OF TWO NEW NUCLEAR GENES, OPSIN AND CAD, WITHIN THE STILETTO FLIES (DIPTERA: THEREVIDAE)

by

HILARY NAN HILL

A thesis submitted to the Graduate Faculty of
North Carolina State University
In partial fulfillment of the
Requirements for the Degree of
Master's of Science

DEPARTMENT OF ENTOMOLOGY

Raleigh

2003

APPROVED BY:

Brian M. Wiegmann **Chair of Advisory Committee**	**Lewis L. Deitz**
Edward L. Vargo	

DEDICATION

This dissertation is dedicated to my parents, Patrick and Carol Hill, who have supported and encouraged all of my choices and endeavors with love and understanding and, to my husband and best friend, Rob Ellison, without whom my life would be incomplete.

BIOGRAPHY

Hilary Hill was born February 29, 1972 in Baltimore, Maryland to Patrick E. Hill and Carol S. Hill. She grew up in Bel Air, MD and upon graduation from Bel Air High School, in 1990, Hilary spent two years at Harford Community College where she received an Associates Degree in Histotechnology. She completed the registry exam to become a board-certified histotechnologist, and started working for Corning® Medical laboratories. After working for two years at the Maryland branch of Corning®, Hilary transferred to the Phoenix, Arizona branch and took a few general courses at Arizona State University.

Two years later Hilary decided to move back to Maryland and enrolled at Towson University (TU) as a full-time biology major while continuing to work as a histotechnologist part-time at the Greater Baltimore Medical Center. While pursuing her degree at TU, she took a course in entomology with Dr. Aubrey Scarbrough. This course spurred Hilary's interest in entomology, and the following semester she enrolled in independent research with Dr. Scarbrough, learning the basic concepts of alpha taxonomy, through studies on robber flies (Diptera: Asilidae). That summer Hilary was selected to participate in the National Science Foundation's Summer Undergraduate Research in Biology Program where she described new species and re-described previously known species of robber flies from Sri Lanka.

After completing her B.Sc., Hilary began her Master's program at North Carolina State University in the fall of 2000 under the supervision of Dr. Brian M. Wiegmann. Her research focused on the investigation of the phylogenetic utility of two new nuclear genes in stiletto flies (Diptera: Therevidae).

Hilary met Rob Ellison in the summer of 1995, before she moved to Arizona. They married in June of 2001, during the second year of Hilary's Master's program.

ACKNOWLEDGMENTS

I sincerely thank Brian M. Wiegmann for his guidance and support throughout my research program. I thank Lewis L. Deitz and Edward L. Vargo for serving on my advisory committee and am grateful to Brian Cassel, Kevin Moulton, and Shaun Winterton for their aid, advice, and expertise in molecular techniques as well as dipteran systematics. Thanks to LongLong Yang for sharing his molecular data and to Dmitry Mukha for sharing his cloning expertise. I thank the entire therevid PEET team, especially: Martin Hauser, Kevin Holston, Mike Irwin, Gail Kampmeir, Donald Webb, and David Yeates. Thank you to Dr. Aubrey Scarbrough for his support and for encouraging me to pursue an advanced degree in entomology. This project was supported by funds from the National Science Foundation (DEB-9977958), North Carolina State University College of Agriculture and Life Sciences, and the North Carolina Agricultural Research Services.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

CHAPTER 2

PHYLOGENETIC UTILITY OF TWO NEW NUCLEAR GENES, OPSIN AND CAD, WITHIN THE STILETTO FLIES (DIPTERA: THEREVIDAE)

CHAPTER 1

Molecular Evolution of Opsin in Therevidae and Their Close Relatives

ABSTRACT

A recent concerted effort is underway to find and increase access to nuclear genes as systematic characters, especially protein-encoding genes that are highly conserved and low copy-number that would resolve phylogenetic divergences from the Mesozoic era, a time of important holometabolous insect diversification. This study investigates the phylogenetic utility of opsin, a new nuclear protein-encoding gene in the dipteran family Therevidae. The present analyses of nucleotide sequence data of opsin reconstructed a therevid phylogeny that is partially concordant with previous classifications. This demonstrates that opsin may have some limited utility as a phylogenetic marker within the Therevidae and closely related Diptera, but also shows that opsin may be better used in combination with other molecular markers or morphological data sets. Phylogenetic analyses of opsin at multiple levels within insects supports these results, but also reveals the significant obstacles in technical manipulation of the gene and in the interpretaion of ortholog/paralog relationships.

INTRODUCTION

Over the past 20 years, technological and theoretical advances such as PCR and automated DNA sequencing has spurred a revolution in molecular biology (Saiki et al. 1985, Mullis et al. 1986, Mullis & Faloona 1987). Easy and repeatable gene amplification using PCR provides systematists a valuable tool for investigating evolutionary relationships. The foundation of molecular systematics is the idea that gene sequences contain phylogenetic information and show a positive correlation of sequence difference and divergence time (Crick 1958, Zuckerkandl & Pauling 1962, Graybeal 1994). Systematists now recognize that the rate of evolution of a gene is closely associated to its utility as a phylogenetic marker (Hillis & Dixon 1991, Graybeal 1994). With this rapid growth in both technology and comparative methods for genetic data, computerized databases (i.e. GenBank, SwissProt, Flybase, etc.) of gene sequences have made numerous gene regions newly accessible for phylogenetic research (Brower & DeSalle 1994, Graybeal 1994).

Until recently the main focus of most systematic research has been on only a small subset of genes, particularly the 18S nuclear ribosomal RNA (rRNA) gene and the genes of the mitochondrial genome (Brower & DeSalle 1994). One reason for the popularity of these genes is that they were the first ones available due to their abundance and the universality of primer regions across taxonomic groups (Brower & DeSalle 1994). Also, RNA sequencing became a reality before DNA sequencing was reasonably tractable (Brower & DeSalle 1994). Mitochondrial genes are often used because they are easily manipulated, abundant, clonally inherited, single-copy, and non-recombinant (Brower & DeSalle 1994, Simon et al. 1994). These genes have proven to be most useful for relatively recent divergences such as those of the mid-Tertiary and later (Kocher et al. 1989, Simon et al. 1994, Lopez et al. 1997) but some have been used for deeper divergences as well (Ballard et al. 1992, Pashley & Ke 1992, Khambamphati 1995, Zardoya & Meyer 1996, Naylor & Brown 1998).

A concerted search is underway to find and increase access to nuclear genes as systematic characters, especially protein-encoding genes that are highly conserved and low copy-number (Friedlander et al. 1992, Brower & DeSalle 1994, Friedlander et al. 1994, Graybeal 1994, Cho et al. 1995, Wiegmann et al. 2000a). Conservation of amino acids make primer design and sequence alignment possible for deeper comparisons, while synonymous substitutions suggest that these genes can also contain information about lower-level taxonomic relationships (Friedlander et al. 1992, Brower & DeSalle 1994, Friedlander et al. 1994, Graybeal 1994, Cho et al. 1995, Wiegmann et al. 2000a). Indeed, nuclear genes have proven to be useful for the more recent Tertiary, as well as for mid-Paleozoic and older divergences (Cammarono et al. 1992, Hasegawa et al. 1993, Kojima et al. 1993, Reiger & Schultz 1997).

There have been few studies of genes that would resolve phylogenetic divergences from the Mesozoic era, a time of important holometabolous insect diversification (Wiegmann et al. 2000a). Our laboratory has been investigating novel phylogenetic markers (specifically nuclear, protein-encoding genes) that have the potential to resolve relationships among brachyceran fly groups with Mesozoic origins. Family-level divergences among the lower brachyceran Diptera are thought to be as old as the Jurassic (150-200 mya) based on known compression fossils (Hennig 1981, Evenhuis 1994, Wiegmann et al. in press), but divergence times within families are more difficult to predict, because therevid fossils are largely unavailable until Baltic amber (Eocene/Oligocene 25-50 mya) (Evenhuis 1994, Grimaldi & Cumming 1999). Consequently, several genes may be necessary to provide enough information to resolve the range of possible ages within such groups. In short, many new genes may need to be sampled.

Opsin: a New Gene for Insect Phylogenetics

The nuclear protein encoding gene opsin has been proposed as one of these new sources of molecular phylogenetic characters for insect divergences below the family level (Mardulyn and Cameron 1999, Cameron & Mardulyn 2001, Ascher et al. 2001, Rokas et al. 2002). Opsin or rhodopsin is a multiple copy, nuclear visual pigment gene that encodes for light-absorbing

proteins, which mediate color vision. Much is known about opsins and their role in color vision, because the system has been studied extensively at both the physiological and biochemical levels and in a variety of organisms (Ovchinnikov 1982, Hargrave et al. 1983, Martynov et al. 1983, Mullen & Akhtar 1983, Nathans & Hogness 1983, Ovchinnikov et al. 1983, Barclay & Findlay 1984, Nathans & Hogness 1984, O'Tousa et al. 1985, Zuker et al. 1985). Most early biochemical studies of visual pigments focused on bovine rhodopsin because of the availability of cattle eyes, and these studies resulted in the determination of the complete amino acid sequence of the protein (Hargrave et al. 1983, Ovchinnikov et al. 1983), as well as its transmembrane topology (Ovchinnikov 1982, Martynov et al. 1983, Mullen & Akhtar 1983, Barclay & Findlay 1984). Other molecular genetic analyses of opsin genes in bovines (Nathans & Hogness 1983), humans (Nathans & Hogness 1984), and *Drosophila* (O'Tousa et al. 1985, Zuker et al. 1985) have facilitated the characterization of opsin genes in various vertebrates and invertebrates.

Visual pigments share a common structural plan, that is, an apoprotein, opsin, is covalently joined to an 11-*cis* retinal-based chromophore through a protonated Schiff's base linkage between the *e*-amino group of a lysine and the aldehyde of retinal (Wald 1968) (Fig. 1). When a photon of light is absorbed, the visual pigment is activated by the isomerization of the 11-*cis* retinal to an all-*trans* configuration (Crandall & Cronin 1997, Chang et al. 1995) (Fig. 2). Photoisomerization triggers a series of conformational changes in the apoprotein resulting in an enzymatically active visual pigment and ultimately producing a neural signal (Crandall & Cronin 1997, Nathans et al. 1986). This is known as visual transduction, the process by which incoming light is converted to neural signals that can be passed to the brain.

The number of visual pigments in concert with their range of spectral absorption determines the visual sensitivity spectrum of an organism (Crandall & Cronin 1997). Opsins are part of the G-protein-coupled receptor family and contain seven transmembrane domains (Briscoe & Chittka 2001). The chromophore sits within a binding pocket formed by the helical transmembrane domains (Briscoe & Chittka 2001). It is within this pocket that specific amino

acids interact to shift the sensitivity of the short-wavelength (377-400 nm)-absorbing chromophore to longer wavelengths of light (Seki & Vogt 1998).

Differences in visual sensitivity may be heavily influenced by environmental conditions (Yokoyama & Yokoyama 1996). Animals inhabit a variety of photic environments, ranging from extreme darkness to bright light. Also, mimicry, protective coloration, and sexual/territorial displays are highly vision-dependent and are only effective when interpreted appropriately by members of the same species or by their predators (Yokoyama & Yokoyama 1996). Therefore, vision can have acute effects on the evolution of organisms by affecting fitness through behaviors such as mating, foraging, and predator avoidance (Yokoyama & Yokoyama 1996). Based on these lines of reasoning, it follows that animal vision has become optimized to fit their respective ecological niches (Yokoyama & Yokoyama 1996).

The number of known visual pigments ranges from one to more than ten and the spectral absorption (λ_{max}) ranges from 350 to 620 nm (Cronin & Marshall 1989, Chang et al. 1995). These diverse ranges in number of visual pigments and spectral absorption are supported by an underlying genetic variation in the visual pigment genes (Chang et al. 1995). This variation in visual pigment genes is evident in insects, where certain insect species such as flies have six kinds of spectral receptors (Hardie 1986), and others such as the owlfly *Ascalaphus macaronius,* have only one (Gogala 1967). The spectral ranges of insects also demonstrate the variation that is seen in other organisms, from the narrow range found in *A. macaronius* (300-480 nm) to the four to six kinds of receptors present in some species of Lepidoptera, Odonata, and Hymenoptera spanning some of the highest visual ranges (<300- >700 nm) known in animals (Gogala 1967, Arikawa et al. 1987, Briscoe 2000, Horridge et al. 1984, Meinertzhagen et al. 1983, Paul et al. 1986, Peitsch et al. 1992, Yang & Osorio 1991).

Until recently, studies of opsin as a phylogenetic marker have been limited to relationships within the Hymenoptera (Mardulyn & Cameron 1999, Cameron & Mardulyn 2001, Ascher et al. 2001, Rokas et al. 2002). Mardulyn & Cameron (1999) used opsin (LW *Rh*) to

investigate relationships of four tribes encompassing the corbiculate bees within the subfamily Apinae (Hymenoptera: Apidae). These workers found that all of the tribal clades, as well as the relationships among the tribes were supported by high bootstrap support values (BSVs) and concluded that LW *Rh* was phylogenetically informative at the tribal and subfamilial levels in Apidae (Mardulyn & Cameron 1999).

Ascher et al. (2001) expanded Mardulyn & Cameron's study (1999) by increasing the taxon sampling. Their study added 52 bee species from 24 tribes and all six extant bee families within Apoidea (Insecta: Hymenoptera). These authors found that opsin (LW *rh*) failed to significantly support the monophyly of corbiculate bees, for relationships among corbiculate bees, or for most other well-established higher-level relationships among long-tongued bees. They did however conclude that opsin provided useful phylogenetic signal at lower taxonomic levels such as genera and tribes (Ascher et al. 2001). When Ascher et al. (2001) combined the expanded opsin dataset with morphological data, the traditional phylogeny of the corbiculate bee tribes was unambiguously supported.

Rokas et al. (2002) assessed the utility of eight DNA sequence markers, including opsin (long-wavelength), in reconstructing phylogenetic relationships at different levels of divergence within the gallwasps (Hymenoptera: Cynipidae). They reported that opsin may be potentially useful at the within family-level divergences (50-100 mya) for gallwasps and other insects (Rokas et al. 2002).

The goal of our study was to apply new molecular data in the form of opsin to test the higher-level therevid phylogeny produced by Yang et al. (2000) and to determine the utility of opsin as a phylogenetic marker within Diptera.

Stiletto Flies (Diptera: Therevidae): Taxonomic Background and Natural History

Therevidae (stiletto flies) have a worldwide distribution, occurring in all geographical regions except Antarctica (Irwin & Lyneborg 1989). Therevids occur in a multitude of habitats including rainforests, coastal dunes, and deserts, with greatest diversity evident in arid

environments where the sandy to sandy-loam soils provide a suitable habitat for the soil-dwelling larvae (Irwin 1976, Winterton et al. 2001).

Therevidae, along with Asilidae, Apioceridae, Apsilocephalidae, Bombyliidae, Mydidae, and Scenopinidae, form the superfamily Asiloidea based on the apomorphic position of the larval posterior spiracles in the penultimate abdominal segment (Hennig 1973, Woodley 1989, Yeates 1994 & 2002). Placing Therevidae among their closest relatives in the Asiloidea, however, has been confounded due to the lack of strong synapomorphies at and above the family level (Nagatomi 1977, Woodley 1989, Yeates 1994 & 2002, Yeates & Wiegmann 1999). Historically, Therevidae were considered to be most closely related to the family Scenopinidae based on the secondary segmentation of the larval abdomen, which gives the appearance of 19 abdominal segments (Irwin & Lyneborg 1981b). Yeates (2002) suggested that Apsilocephalidae is the sister-group to Therevidae based on the presence of a distinct knob on the anterior surface of the hind coxa, a characteristic not found in scenopinids. Interestingly, the larvae of apsilocephalids are unknown and thus, it is not known whether the larvae have secondary segmentation (Yeates 2002). Recent research, based on both morphological and molecular data, suggests that the Scenopinidae, Apsilocephalidae, and Therevidae form a well-supported monophyletic group among the asiloid families (Yang 2000, Yeates 2002, Yeates et al. submitted), but the exact relationship among them remains uncertain (Yeates et al. submitted).

Until the recent advent of the Partnerships for Enhancing Expertise in Taxonomy (PEET) therevid research program (Irwin & Kampmeier 1997), stiletto flies were relatively understudied, with no phylogeny-based classification for the worldwide taxa (Yang et al. 2000). Recent collections in poorly collected regions such as Australia, Patagonia, and Southern Africa have enabled the discovery and description of many new genera and species of Therevidae (Irwin & Yeates 1998, Winterton & Irwin 1999, Winterton et al. 1999, Winterton et al. 2000, Winterton et al. 2001, Yeates et al. submitted). The monophyly of Therevidae was never firmly established by morphological evidence (Woodley 1989, Yeates 1994), which was postulated by Yang et al.

(2000) to be due to the difficulty of placing several enigmatic groups sometimes included within Therevidae. Approximately 1000 species of therevids have been described thus far, and this is likely a fraction of the actual therevid diversity (Irwin, pers. comm.).

Species of Therevidae are presently divided into three subfamilies: Phycinae, Therevinae (Lyneborg 1976, Irwin & Lyneborg 1981a) and Agapophytinae (Winterton et al. 2001). Phycinae has been divided into two tribes, Phycini and Xestomyzini (Lyneborg 1976), Therevinae includes the tribe Cyclotelini (Gaimari & Irwin 2000), but a tribal-level classification has not been proposed for Agapophytinae.

The molecular phylogenetic analyses of Yang et al. (2000) and Yang (2000) were the first studies to apply nucleotide sequence data (EF-1α and 28S rDNA) to estimate relationships within Therevidae (Fig. 3). Yang's studies resulted in several important discoveries: First, the Therevidae are monophyletic, but does not include Apsilocephala; second, the earliest therevid divergences divide the family into two major monophyletic clades that correspond closely with Irwin and Lyneborg's (1981a) division of Therevidae into Phycinae and Therevinae, and third, a large, diverse radiation of the Australasian taxa is placed at the base of the Therevinae. Nonetheless, the molecular data sets of Yang et al. (2000) and Yang (2000) held insufficient variation to conclusively specify relationships within and among lineages of Therevinae.

In this study, we investigate opsin's utility within Therevidae through multiple phylogenetic experiments. We analyze multiple levels in the gene-tree/species-tree hierarchy for insect opsins due to the complexity of interpreting the history of a multiple copy gene. The first experiment includes only published insect opsin sequences in order to interpret ortholog/paralog relationships. In the second experiment, we add 42 sequences from Therevidae and their close relatives to investigate ortholog/paralog relationships for our newly characterized opsins relative to established sequences. Next, we restricted our dataset to include only dipteran opsin sequences. The aim of this experiment was to further investigate ortholog/paralog relationships when limited to a single monophyletic insect order (Diptera). The fourth experiment used

sequence only from Therevidae and their close relatives. This experiment was then subdivided into analyses including a long wavelength paralog (copy) of opsin ("lw" taxa) and an analysis of only a short wavelength ortholog ("ops" taxa). The main objective of the final experiment was to determine this short wavelength opsin's utility for phylogeny estimation within Therevidae.

Our analyses of opsin acheived the following: 1) we have produced 42 new sequences of opsin in Diptera; 2) we have analyzed our data in the context of published insect opsin sequences; 3) we have conducted the first analysis of opsin's phylogenetic utility in Diptera; 4) we have discovered multiple copies of opsin in Therevidae; 5) we have provided an additional molecular dataset for therevids; 6) we have reconstructed a therevid phylogeny that is partially concordant with previous classifications; and 7) we have discovered some of the practical difficulties and analytical complexities involved in phylogenetic research using opsin.

MATERIALS AND METHODS

Taxon Sampling

We obtained nucleotide sequences from a total of 23 genera of Therevidae, including the three recognized subfamilies, Phycinae and Therevinae (Lyneborg 1976; Irwin & Lyneborg, 1981a) and Agapophytinae (Winterton et al. 2001) (Table 1). Two Asiloid genera from the putative sister-group Apsilocephalidae, as well as from the closely related Scenopinidae were sampled as outgroups. An additional two asiloid genera from Bombyliidae were sequenced, as well as two genera from Empididae. All specimens were collected directly into 95% ethanol and stored at -80ºC in the laboratory. All other insect sequences used were retrieved from GenBank (Table 2).

Laboratory Methods

Total genomic nucleic acids were extracted using the guanidinium isothiocyanate homogenization protocol of Chirgwin et al. (1979). Nucleic acid pellets were resuspended in 50µl TE buffer and stored at -80˚C. Whole specimens or partial specimens (i.e.: heads, wings,

and legs) were stored in 95% ethanol at the Department of Entomology of North Carolina State University Insect Collection Genome Bank.

Oligonucleotide primers (Table 3) designed to amplify DNA were synthesized by Genosys Biotechnologies, Inc. (Woodlands, Texas). We initially amplified an approximately 650 bp fragment of opsin for most taxa using degenerate primers (ops102f/ops336r) based on an alignment of published dipteran opsin sequences downloaded from GenBank and one newly characterized therevid opsin sequence. The majority of products from the initial amplification of opsin were then re-amplified using specific primers (ops1f/ops3r) designed based on the available therevid sequence. This amplification resulted in an approximately 350bp fragment. In some cases, fragments were re-amplified with primer pair ops1f/ops336r to get a larger fragment (~700bp). Some taxa (denoted by "lw" after their name) were amplified using primers (lwrhf/lwrhr) developed by Donat Agosti (AMNH) based on published opsin sequences of *Drosophila pseudoobscura* (Carulli & Hartl 1992). PCR products from the initial amplification were then re-amplified using the same primers with M13 tails. All PCR protocols used TaKaRa Ex Taq™ Hot Start Version (TaKaRa Biomedicals, Japan).

Amplification with ops102f/ops336r followed a touchdown polymerase chain reaction (PCR) protocol as follows: initial 94˚C for 4 min followed by 5 cycles of 94˚C for 30 sec, 48˚C for 30 sec, 72˚C for 1 min and then 37 cycles of 94˚C for 30 sec, 43˚C for 30 sec, 72˚C for 1 min followed by an additional elongation step of 72˚C for 3 min. Amplification with lwrhf/lwrhr followed the same touchdown PCR protocol that was used with ops102f/ops336r except for annealing temperatures: 45ºC and 40ºC, respectively. Amplified products (2 lanes of 20µl per taxon) were visualized and cut from a 1% low melt gel, by adding 100µl of depc water to the gel slices and then melting it in a 70ºC water bath for 2 min. Gel cuts were re-amplified using the following PCR protocol: initial 94˚C for 5 min followed by 36 cycles of 94ºC for 1 min, 42ºC for 1 min, 72ºC for 1 min followed by an additional elongation step of 72ºC for 5 min. Products

from re-amplification (1 lane of 25µl per taxon) were then visualized and cut from a 1% low melt gel and purified using the Qiaquick® Gel Extraction Kit (Qiagen, Inc., Santa Clara, CA).

Purified products were directly sequenced using d-Rhodamine Terminator Cycle Sequencing Ready Reaction with AmpliTaqFS® DNA polymerase (PE Applied Biosystems, Foster City, CA) following the suggested protocol except annealing temperature used was 40ºC instead of the suggested 50ºC for improved results. Cycle sequencing reactions used the same primers that were used in the re-amplification PCR protocols listed in Table 3 except for those taxa that were re-amplified using M13 tailed primers (Table 3). We sequenced sense and antisense DNA strands for confirmation of all sequences. Sequences were gel fractionated and base-called on an ABI PRISM 377 DNA automated sequencer (PE Applied Biosystems, Foster City, CA).

Nucleotide Alignment and Phylogenetic Analysis

Therevid sequence editing, contig assembly, and consensus calculations were performed in Sequencher 4.1 (Gene Codes Corp., Ann Arbor, MI) for the Power Macintosh. Consensus sequences were aligned manually using the multiple alignment editor of Genetic Data Environment (GDE 2.2, Smith et al. 1994). Gaps in aligned regions were treated as missing data. Introns were identified by comparing therevid opsin sequences to known insect opsin sequences (Carulli et al. 1994), and were excluded from the data set prior to phylogenetic analysis due to their considerable sequence divergence and resulting alignment instability. Introns from the insect nucleotide sequences downloaded from GenBank were manually excised and the remaining sequences were translated into their component amino acids. All translated sequences were then checked for accuracy against the GenBank amino acid translations. Manual alignment of the nucleotides and their corresponding amino acids of both therevids and GenBank sequences were based on the published alignment of Carulli et al. (1994).

Equally Weighted Parsimony. Maximum parsimony (MP) analyses were implemented in PAUP* version 4.0 (Swofford 2002) using the heuristic search option and 20 replicate random

addition searches with tree-bisection reconnection (TBR) branch swapping. In all MP analyses the following settings were used: all characters were treated as unordered, all characters have equal weight, gaps are treated as missing, multistate taxa treated as uncertainty, steepest descent option not in effect, MaxTrees unlimited (auto-increased by 100), branches collapsed if maximum branch-length is zero, and MulTrees option in effect. When heuristic searches resulted in multiple equally parsimonious trees (EPTs), we summarized them in a strict consensus tree. Branch support estimates for most parsimonious trees (MPTs) and EPTs were calculated with nonparametric bootstrap analysis (Felsenstein 1985). Bootstrap analyses were performed with 1000 pseudoreplications of random addition sequences and TBR branch swapping.

Neighbor-Joining. Neighbor-joining (NJ) methods performed in PAUP* 4.0 were implemented when MP methods were computationally infeasible. In all NJ analyses the following settings were used: optimality criterion equals distance (minimum evolution); negative branch lengths allowed, but set to zero for tree-score calculation, ties (if encountered) will be broken systematically; distance measure equals uncorrected ("p"); starting trees obtained via neighbor-joining; branch swapping algorithm equals TBR; steepest descent option not in effect; MaxTrees unlimited (auto-increased by 100); branches not collapsed if maximum branch-length is zero; MulTrees option in effect. Branch support estimates for NJ trees were calculated with nonparametric bootstrap analysis (Felsenstein 1985). Bootstrap analyses were performed with 100 replicates and TBR branch swapping.

RESULTS

Phylogenetic Analyses

Published Insect Opsin Sequences. The MP analysis of all nucleotides of opsin sequence resulted in a single MPT (Fig. 4) of tree length 2201 with a consistency index (CI) of 0.348 and a retention index (RI) of 0.505 (Table 4). The single MPT shows clustering of the different copies of opsin that is similar to the results found by Townson et al. (1998) and the more recent work of Mardulyn & Cameron (1999). Moreover, our study has increased the taxon sampling due to the

increased availability of opsin sequences from more recent works (Chase et al. 1997, Kitamoto et al. 1998, Arca et al. 1999, and Gao et al. 2000).

The presence of multiple orthologous copies is clearly demonstrated by the separate groupings of the ant species (*Camponotus abdominalis* and *Cataglyphis bombycinus*) ops1 and 2. These groupings are concordant with those seen in Mardulyn & Cameron (1999), with *Apis mellifera* lw (green) forming the sister-group to the ant ops1 sequences and the *A. mellifera* uv sequence sister to the ant ops2 sequences. With the addition of the hemipteran sequences, *M. viciae* (lwl and uv), there is further support of orthologous relationships among copies. The *Drosophila* species (*D. melanogaster* and *D. pseudoobscura*) form several clusters of orthologous pairs throughout the tree. However, *D. melanogaster* rh1 does not form a sister-group with *D. pseudoobscura* rh1, which is not concordant with Mardulyn & Cameron (1999). The branch length for *D. melanogaster* rh1 is long in comparison with most of the other observed branch lengths, suggesting that this sequence may be highly unique and divergent in sequence from other *Drosophila* opsins, a property that can greatly affect phylogenetic interpretation (Felsenstein 1978, Swofford et al. 1996, Friedrich & Tautz 1997, Hwang et al. 1998).

Analysis of the same dataset with third codon positions removed also resulted in a single MPT, 1027 steps long with a CI of 0.462 and a RI of 0.645. Topologically, the two trees are very similar with only slight rearrangements of taxa and similar bootstrap support values (BSVs) (tree not shown).

Published Insect Opsin Sequences + Therevidae and Close Relatives. Parsimony analysis of the insect opsin sequences with the addition of 35 therevid taxa and seven close relatives resulted in 15 EPTs with a tree length of 3240, CI of 0.241, and RI of 0.550 (Table 4). A strict consensus of the 15 EPTs is shown in Figure 5.

This tree shows a paraphyletic Therevidae, with *Ataenogera* sp., *A. abdominalis*, *Phycus*, *Pherocera*, *Ruppellia multisetosis*, and *Hoplosathe frauenfeldi* lw forming a clade separate from all other therevids. Of these taxa, all except for *H. frauenfeldi* lw are part of the therevid

subfamily Phycinae, the basal-most subfamily in Therevidae (Yang et al. 2000). The only phycine taxa that fell into the therevid clade were *Hemigephyra* and *Henicomyia* (Xestomyzini), and *Neotabuda.* All non-therevid asiloid taxa form a clade that also includes some of the other insect opsins. *Stenomphrale teutankhameni* (Scenopinidae) opsin joins *D. pseudoobscura* (67% bootstrap support) opsin and together, they are a sister-group to the *Calliphora vicina* opsin gene (90% bootstrap support). This strong support for clustering genes from multiple dipteran lineages suggests that these sequences may represent orthologous copies of opsin that shared a common genic ancestor before the divergence of higher flies.

Our results indicate that we have sequenced two different copies of opsin (paralogs). Strong support for this result is shown by parallel amplifications from the therevid, *H. frauenfeldi,* using both "lw " and "ops" primers (Table 3). The *H. frauenfeldi* sequences do not form a sister-group and are placed in separate clades within the phylogram. Furthermore, the amino acid sequences for the two copies of opsin in *H. frauenfeldi* also differ significantly (42%). Finally, taxa amplified and sequenced using the "lw" primers lacked a conserved intron that we found in all other sequences amplified and sequenced using "ops" primers.

Neighbor-joining (NJ) analysis of these data with third positions removed gave the tree shown in Fig. 6. This tree shows a monophyly for all sampled therevid opsins (86% BSV) with the exception of *H. frauenfeldi* lw, which as discussed above, groups with all other sampled genes that are of the "lw" form.

Dipteran Opsin. The MP analysis of dipteran opsin sequences resulted in 25 EPTs with a tree length of 1865, CI of 0.362, and RI of 0.543 (Table 4). A strict consensus of the 25 EPTs is shown in Fig. 7. This tree shows Therevidae paraphyletic (BSV < 50%), with the same phycine species mentioned in the previous section forming a clade with other dipteran opsin sequences and *H. frauenfeldi* lw clustering with other "lw" taxa. Also, *S. teutankhameni* (Scenopinidae) ("ops" primers) is again a sister-group to *D. pseudoobscura* and *Calliphora vicina* (91% BSV) and this clade is the sister-group to all of the Therevidae and closely related Diptera sequenced

for this study. Expected phylogenetic relationships for opsins sampled from the basal-most therevid group (Phycinae) and the closely related asiloid flies seems to have been confounded by the inclusion of multiple copies from more distantly related dipterans.

Neighbor-joining (NJ) analyses excluding third positions results in the tree shown in Fig. 8. The NJ tree depicts a monophyletic Therevidae (excluding *H. frauenfeldi* lw) with 89% BSV. This tree further supports the presence of two separate copies of opsin within our study group as seen in the clustering of the "lw" taxa.

Therevidae + Their Close Relatives ("lw" Copies Included). The MP analysis of Therevidae and closely related taxa including "lw" copies yielded two EPTs of length 1045, CI = 0.365, and RI = 0.543 (Table 4). Fig.ure 9 is a strict consensus of the two EPTs, which shows Therevidae as a paraphyletic group (70% BSV). *H. frauenfeldi* lw again appears to be more closely related to outgroup "lw" sequences. *Mythicomyia* (Bombyliidae) and *S. teutankhameni* (Scenopinidae) sequences are placed within the therevid clade. These taxa are the only outgroup taxa that were sequenced with the "ops" primers. Therefore, the non-"lw" taxa form a monophyletic clade with respect to the "lw" taxa, which further supports our inference that these represent two separate paralogous copies of opsin.

This analysis loosely supports the traditional classification of Therevidae into the three subfamilies: Phycinae, Therevinae (Lyneborg 1976, Irwin & Lyneborg 1981a) and Agapophytinae (Winterton et al. 2001). The majority of the phycine taxa are correctly placed basal to the rest of the Therevidae, with the exception of *Neotabuda*, which falls into the agapophytine clade and appears as a sister-group to *Agapophytus adonis* (99% BSV). The phycines *Hemigephyra* and *Henicomyia*, which are part of the Xestomyzini tribe, form a clade separate from the rest of the Phycinae with 66% BSV. The Agapophytinae form a cluster (< 50% BSV) with only New Genus S falling outside. The Therevinae also form a cluster separate from the other subfamilies. Thus the general topology of this tree is consistent with previous therevid

classifications (Fig. 3), but has low bootstrap support values and a few taxa fall outside of their expected position.

NJ analysis of just positions 1+2 yields strong support (BSV 87%) for a monophyletic Therevidae (excluding *H. frauenfeldi* lw) (Fig. 10). Overall bootstrap support is low throughout most of the tree. Most of the taxa comprising the Phycinae form a basal, paraphyletic cluster, with the Xestomyzini taxa (*Hemigephyra* and *Henicomyia*) forming a monophyletic clade. There are two separate paraphyletic clusters of agapophytine taxa and the Therevinae are paraphyletic as well. Most of the congeneric taxa also fail to form monophyletic clades. Exclusion of third positions does not help to resolve the traditional subfamilial classification of Therevidae nor does it aid in the reconstruction of lower-level therevid relationships.

Therevidae + Their Close Relatives ("lw" Copies Excluded). The MP analysis of Therevidae with closely related taxa resulted in four EPTs with 718 steps, CI = 0.373, and RI = 0.528 (Table 4). A strict consensus of the four EPTs (Fig. 11) shows a monophyletic Therevidae (BSV = 79%). Again, this analysis loosely supports the traditional classification of Therevidae into the three subfamilies: Phycinae, Therevinae (Lyneborg 1976, Irwin & Lyneborg 1981a) and Agapophytinae (Winterton et al. 2001) with low bootstrap support overall. The topology is very similar to the previous analysis of the same data including "lw" copies, with major clusters representing each of the subfamilies and a few taxa falling outside of the traditional classification.

At the tribal and genus level, opsin appears to retrieve some of the expected relationships, for example the tribe Xestomyzini (*Hemigephyra* + *Henicomyia*), and higher-therevines *(Anabarhynchus* spp.*, Brachylinga* spp.*, Cyclotelus* spp.*, + Ozodiceromia spp).* However some taxa (*Neotabuda* and New genus S) are placed outside their respective subfamilies.

The NJ tree excluding third positions (Fig. 12) supports a monophyletic Therevidae (BSV = 96%) with a paraphyletic Phycinae basal to the rest of the subfamilies. The Xestomyzini (*Hemigephyra* + *Henicomyia*) forms a separate clade from the Phycinae. Again, the exclusion of third positions causes the tree topology to be less concordant with the traditional subfamilial

classification of Therevidae and it also lessens the overall bootstrap support. Some of the genus-level clades (*Anabarhynchus* spp., *Brachylinga* spp., *Cyclotelus* spp.) that were well supported in the analysis including all nucleotide positions are now broken up as well. Exclusion of third positions in this small fragment of opsin appears to severely decrease the phylogenetic signal necessary to resolve relationships within Therevidae.

DISCUSSION

Our analyses of nucleotide sequence data of opsin yields several important results: 1) we have produced 42 new sequences of opsin in Diptera; 2) we have analyzed our data in the context of published insect opsin sequences; 3) we have demonstrated that the gene regions investigated here provide limited resolution for both higher-level and intrafamilial relationships in brachyceran Diptera; 4) we have discovered multiple copies of opsin in Therevidae; 5) we have reconstructed a therevid phylogeny that is partially concordant with previous classifications; 6) we have provided an additional molecular dataset for therevid phylogenetics; 7) we have described some of the practical difficulties and analytical complexities involved in phylogenetic research using opsin.

Historically, the availability of dipteran opsin sequences has been limited to a few taxonomic groups (*Drosophila*, *Calliphora*, and *Anopheles*). Our study has established a substantial dataset of dipteran opsin sequences (42) from other families of flies that can be used by insect molecular systematists for various analyses. This dataset includes 23 genera of Therevidae, including the three recognized subfamilies, Phycinae and Therevinae (Lyneborg 1976; Irwin & Lyneborg, 1981a) and Agapophytinae (Winterton et al. 2001) (Table 1); Asiloid genera from Apsilocephalidae, Scenopinidae, and Bombyliidae; as well as two genera from Empididae. Additionally, our dataset, when used in conjunction with previously published insect opsin sequences, could be a useful tool for primer design to extend the sample within Diptera and to other insect orders.

On a broader scale, published insect opsin sequences have been characterized in six insect orders: Diptera, Hymenoptera, Lepidoptera, Orthoptera, Mantodea, and Hemiptera. As part of our study, we have conducted the most comprehensive phylogenetic analysis of published insect opsins to date, increasing the dataset by an additional 42 dipteran opsin sequences and also adding several newly characterized insect opsins from other studies (Chase et al. 1997, Kitamoto et al. 1998, Arca et al. 1999, and Gao et al. 2000). This analysis included all of the characterized copies of opsin from each order and when possible, different families, genera, and species representatives from these orders. Our phylogenetic reconstruction clearly demonstrates the presence of multiple copies of opsin through the clustering of orthologs across different species. These results agree with those of Townson et al. (1998) and the more recent work of Mardulyn & Cameron (1999) for less comprehensive samples. Furthermore, opsin gene trees reveal that at least two copies of opsin are present in lower brachyceran Diptera.

As mentioned in the Materials and Methods section, we used two sets of primers to sequence opsin: the "ops" primers, which were used with the majority of the therevid taxa and two of the outgroups, and the "lw" primers, which were used on the majority of the outgroup taxa and one therevid. Our successful amplification and sequencing using both sets of primers from the same specimen of the therevid *Hoplosathe frauenfeldi* is direct evidence for the presence of multiple copies of opsin in lower brachyceran flies. On editing these sequences, we discovered the presence of an intron in all of the "ops"-primer generated sequences. This intron is absent in the amplifications using the "Lwrhf/r" primers. The two paralogs from *H. frauenfeldi* also had significant differences in amino acid sequence (42%). Phylogenetic analysis of all published insect opsins show that the two sequences of *H. frauenfeldi* did not form a clade. The same result was also found when we analyzed only dipteran opsin sequences. In general all analyses that included the "ops" amplified fragment yielded clustering of these sequences consistent with the interpretation that these are orthologs, having split by gene duplication from the "lw" copy prior to the diversification of the lower brachyceran lineages.

Most previous studies of dipteran opsin sequences have been limited to the study of gene evolution (O'Tousa et al. 1985, Cowman et al. 1986, Montell et al. 1987, Zuker et al. 1987, Huber et al. 1990, Carulli and Hartl 1992, Carulli et al. 1994, Chou et al. 1996, Crandall & Cronin 1997, Townson et al. 1998, Arca et al. 1999). To date, studies of opsin's utility as an organismal phylogenetic marker have only been carried out within the Hymenoptera (Mardulyn & Cameron 1999, Cameron & Mardulyn 2001, Ascher et al. 2001, Rokas et al. 2002).

Assessing opsin as a marker for dipteran phylogeny is confounded by the presence of multiple copies. Both "lw" and "ops" sequences have potential utility because they are obtainable through direct amplification from multiple dipteran families. The fragments investigated here (350bp and 700bp, respectively) are short relative to most commonly used genes in insect systematics (Caterino et al. 2001), and will be limited by constraints on protein evolution and subject to homoplasy in third positions. Nonetheless, our data show that significant variation can be found in all three nucleotide positions and that these sites provide some evidence for relationships at and below the family level.

Mardulyn & Cameron (1999) found that opsin was useful for resolving phylogenetic relationships at the subfamilial and tribal levels in corbiculate bees, with divergence times as old as the Cretaceous (50-150 mya). Additional analysis of an expanded dataset by Ascher et al. (2001) found that opsin (LW *rh*) also successfully resolved bee relationships at lower taxonomic levels (within tribes and genera). More recently, Rokas et al. (2002) assessed opsins (long-wavelength) utility in reconstructing phylogenetic relationships at different levels of divergence within the gallwasps (Hymenoptera: Cynipidae). They reported that opsin may be potentially useful at the within family-level divergences (50-100 mya) for gallwasps and other insects (Rokas et al. 2002). Our analyses of opsin in Therevidae are similarly promising. Family-level divergences among the lower brachyceran Diptera are hypothesized to be as old as the Jurassic (150-200 mya) based on known compression fossils (Hennig 1981, Evenhuis 1994), but divergence times within Therevidae and others could be as recent as Eocene/Oligocene (25-50

mya) (Baltic amber; Evenhuis 1994, Grimaldi & Cumming 1999). Although our opsin-based therevid phylogeny was not entirely concordant with previous classifications (Yang 2000; Yang et al. 2000), and BSVs were generally low, clusters of taxa fell within the expected subfamily boundaries, and with only a few exceptions. Also, many of the tribal and generic relationships were concordant with previous classifications. Our results suggest that opsin could be useful at divergence times between 50-100mya, although further analysis with a larger fragment would more conclusively support this finding.

Finally, our study has yielded significant insights into the practical and technical considerations of using opsin as a phylogenetic marker. Due to opsin's limited use in previous phylogenetic studies and, it being part of a multiple copy gene family, amplifying and sequencing a single copy of it was extraordinarily difficult. Primers that were used in previous studies of opsin in insects were initially tried with limited -to-no success. Consequently, it was necessary to design therevid-specific primers to amplify and sequence opsin.

The primers that worked in the majority of therevid taxa only amplified an approximately 350 bp fragment, a relatively small fragment to use for phylogenetic analyses. Due to their specificity to therevids, we were only able to amplify and sequence two of our outgroup taxa (*Mythicomyia* and *S. teutankhameni*). In order to sequence more of our outgroup taxa, we used the Lwrhf/Lwrhr primers designed to include standard M13 tails (see Materials and Methods section). We also tried using these altered primers with many of our ingroup taxa and only a single taxon was successfully sequenced. The difficulties associated with amplification and sequencing a single ortholog may seriously limit opsin's potential as a phylogenetic marker in other taxa. This gene required considerable time and effort for exploratory amplification and sequencing and multiple rounds of primer design. The tradeoff between incorporating necessary ingroup primer specificity and also sufficient generality (=primer degeneracy) to obtain outgroup sequences may limit the use of opsin in groups that are as diverse and old as Therevidae and

related lower Brachycera. Its ultimate use may be limited to recently diverged taxa for which primers will be less variable, or as a gene fragment used only in combination with other genes.

CONCLUSIONS

Opsin is an important potential source of phylogenetic information in Therevidae and closely related Diptera. Molecular systematic analysis of opsin at multiple levels within insects supports this assertion, but also reveals significant hurdles in technical manipulation of the gene and in the interpretation of ortholog/paralog relationships. While we have significantly increased the sample of opsin genes now available, our study suggests that these may be most useful in combination with other genes.

In light of our results, there are several opportunities for future analyses using opsin within Therevidae. Future work with opsin in therevids should include a characterization and analysis of the phylogenetic utility of the other copies of opsin. Also, increased taxon sampling, with better representation from all three subfamilies of Therevidae would provide a broader perspective on the gene's phylogenetic value. The opsin data generated from this study should be added to the existing therevid molecular dataset (Yang et al. 2000) to determine whether it helps resolve relationships when used in combination with other molecular data. Ultimately, therevid opsin should be compared to, and combined with, important comparative morphological evidence now emerging from ongoing detailed studies (Metz 2002, Holston in preparation).

ACKNOWLEDGMENTS

We thank L.L. Deitz and E.L. Vargo for comments on this manuscript. We are indebted to B. Cassel and K. Moulton for their technical support and assistance. The following people provided important specimens, data, and/or information that enabled us to conduct this research: M. Hauser, K. Holston, M. Irwin, G. Kampmeir, K. Moulton, E. Schlinger, S. Winterton, L. Yang, and D. Yeates. This project was supported by funds from the NSF/PEET grant (DEB-9977958), the North Carolina State University College of Agriculture and Life Sciences, and North Carolina Agricultural Research Services.

REFERENCES

Arca, B., F. Lombardo, M. de Lara Capurro, A. della Torre, G. Dimopoulos, A.A. James, M. Coluzzi. 1999. Trapping cDNAs encoding secreted proteins from the salivary glands of the malaria vector *Anopheles gambiae*. Proc. Natl. Acad. Sci. USA. 96: 1516-1521.

Arikawa, K., K. Inokuma, E. Eguchi. 1987. Pentachromatic visual system in a butterfly. Naturwissenschaften 74: 297-298.

Ascher, J.S., B.N. Danforth, J. Shuqing. 2001. Phylogenetic utility of the major opsin in bees (Hymenoptera: Apoidea): a reassessment. Mol. Phylogenet. Evol. 19: 76-93.

Ballard, J.W.O., G.J. Olsen, D.P. Faith, W.A. Odgers, D.M. Rowell, P.W. Atkinson. 1992. Evidence from 12S ribosomal RNA that onychophorans are modified arthropods. Science 258: 1345-1348.

Barclay, P.L., J.B.C. Findlay. 1984. Labeling of the cytoplasmic domains of ovine rhodopsin with hydrophilic chemical probes. Biochem. J. 220: 75-84.

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., L. Chittka. 2001. The evolution of color vision in insects. Annu. Rev. Entomol. 46: 471-510.

Brower, A.V.Z., R. DeSalle. 1994. Practical and theoretical considerations for choice of a DNA sequence region in insect molecular systematics, with a short review of published studies using nuclear gene regions. Ann. Entomol. Soc. Am. 87: 702-716.

Cameron, S.A., P. Mardulyn. 2001. Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera: Apinae). Syst. Biol. 50: 194-214.

Cammarono, P., P. Palm, R. Creti, E. Ceccarelli, A.M. Sanangelantoni, O. Tiboni. 1992. Early evolutionary relationships among known life forms inferred from elongation factor EF-2/EF-G sequences: Phylogenetic coherence and structure of the Archaeal domain. J. Mol. Evol. 34: 396-405.

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

Carulli, J.P., D.-M. Chen, W.S. Stark, D.L. Hartl. 1994. Phylogeny and physiology of *Drosophila* opsins. J. Mol. Evol. 38: 250-262.

Caterino, M.S., S. Cho, F.A.H. Sperling. 2000. The current state of insect molecular systematics: a thriving tower of Babel. Annu. Rev. Entomol. 45: 1-54.

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

Chang, B.S.W., D. Ayers, W.C. Smith, N.E. Pierce. 1996. Cloning of the gene encoding honeybee long-wavelength rhodopsin: a new class of insect visual pigments. Gene 173: 215-219.

Chase, M.R., R.R. Bennett, R.H. White. 1997. Three opsin-encoding cDNAs from the compound eye of *Manduca sexta*. J. Exp. Biol. 200: 2469-2478.

Chirgwin, J.M., A.E. Przybyla, R.J. MacDonald, W.J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18: 5294.

Cho, S., A. Mitchell, J.C. Reiger, C. Mitter, R.W. Poole, T.P. Friedlander, S. Zhao. 1995. A highly conserved nuclear gene for low-level phylogenetics: elongation-factor 1α recovers morphology-based tree for heliothine moths. Mol. Biol. Evol. 12: 650-656.

Chou, W.-H., K.J. Hall, D.B. Wilson, C.V. Wideman, S.M. Townson, L.V. Chadwell, S.G. Britt. 1996. Identification of a novel *Drosophila* opsin reveals specific patterning of the R7 and R8 photoreceptor cells. Neuron 17: 1101-1115.

Cowman, A.F., C.S. Zuker, G.M. Rubin. 1986. An opsin gene expressed in only one photoreceptor cell type of the *Drosophila* eye. Cell 44: 705-710.

Crandall, K.A., T.W. Cronin. 1997. The molecular evolution of visual pigments of freshwater crayfishes (Decapoda: Cambaridae). J. Mol. Evol. 45: 524-534.

Crick, F.H.C. 1958. On protein synthesis. Symp. Soc. Exp. Biol. 12: 138-163.

Cronin, T.W., N.J. Marshall. 1989. A retina with at least ten spectral types of photoreceptors in a mantis shrimp. Nature 339: 137-140.

Evenhuis, N.L. 1994. Catalogue of the Fossil Flies of the World (Insecta: Diptera). Backhuys, Leiden.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27: 401-410.

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

Flook, P.K., C.H.F. Rowell. 1998. Inferences about orthopteroid phylogeny and molecular evolution from small subunit nuclear ribosomal DNA sequences. Insect Mol. Biol. 7: 163-178.

Friedlander, T.P, J.C. Reiger, C. Mitter. 1992. Nuclear gene sequences for higher level phylogenetic analysis: 14 promising candidates. Syst. Biol. 41: 483-490.

Friedlander, T.P., J.C. Reiger, C. Mitter. 1994. Phylogenetic information content of five nuclear gene sequences in animals: Initial assessment of character sets from concordance and divergence studies. Syst. Biol. 43: 511-525.

Friedrich, M., D. Tautz. 1997. An episodic change of rDNA nucleotide substitution rate has occurred at the time of the emergence of the insect order Diptera. Mol. Biol. Evol. 14:644-653.

Gaimari, S.D., & M.E. Irwin. 2000. Phylogeny, classification, and biogeography of the cycloteline Therevinae (Diptera: Therevidae). Zool. J. Linn. Soc. 129: 129-240.

Gao, N., R.G. Foster, J. Hardie. 2000. Two opsin genes from the vetch aphid, *Megoura viciae*. Insect Mol. Biol. 9: 197-202.

Gogala, M. 1967. Die spektrale empfindlichkeit der doppelaugen von *Ascalaphus macaronius* Scop. (Neuroptera, Ascalaphidae). Z. Vergl. Physiol. 57: 232-243.

Graybeal, A. 1994. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. Syst. Biol. 43: 174-193.

Grimaldi, D., J. Cumming. 1999. Brachyceran Diptera in Cretaceous ambers and Mesozoic diversification of the Eremoneura. Bull. Am. Mus. Nat. Hist. 239: 1-124.

Hardie, R.C. 1986. The photoreceptor array of the dipteran retina. Trends Neurosci. 9: 419-423.

Hargrave, P.A., J.H. McDowell, D.R. Curtis, J.K. Wang, E. Juszczak, S.L. Fong, J.K. Mohanna Rao, P. Argos. 1983. The structure of bovine rhodopsin. Biophys. Struct. Mech. 9: 235-244.

Hasegawa, M., T. Hashimoto, J. Adachi, N. Iwabe, T. Miyata. 1993. Early branchings in the evolution of eukaryotes: ancient divergence of *Entamoeba* that lacks mitochondria revealed by protein sequence data. J. Mol. Evol. 36: 380-388.

Hennig, W. 1973. Diptera. In "Handbuch der Zoologie, IV: Arthropoda" (W. Kukenthal, Ed.), pp. 1-337. de Gruyter, New York.

Hennig, W. 1981. Insect Phylogeny. Wiley, New York.

Hillis, D.M., M.T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. 66: 411-453.

Holston, K.C. A phylogenetic assessment of *Thereva* Latreille (Insecta: Diptera: Therevidae). Ph.D. Thesis. Department of Entomology, University of Illinois at Urbana-Champaign, in preparation.

Horridge, G.A., L. Marcelja, R. Jahnke. 1984. Colour vision in butterflies. I. Single colour experiments. J. Comp. Physiol. A 155: 529-542.

Huber, A., D.P. Smith, C.S. Zuker, R. Paulsen. 1990. Opsin of *Calliphora* peripheral photoreceptors R1-6. Homology with *Drosophila* Rh1 and posttranslational processing. J. Biol. Chem. 265: 17906-17910.

Hwang, U.W., W. Kim, D. Tautz, M. Friedrich. 1998. Molecular phylogenetics at the Felsenstein zone: approaching the Strepsiptera problem using 5.8S and 28S rDNA sequences. Mol. Phylogenet. Evol. 9: 470-480.

Irwin, M.E. 1976. Morphology of the female terminalia and known ovipositing behaviour of female Therevidae (Diptera: Asiloidea), with an account of correlated adaptations and comments on phylogenetic relationships. Ann. Natal Mus. 22: 913-935.

Irwin, M.E., G. Kampmeier. 1997. PEET: a training effort that is paying dividends. Ill. Nat. Hist. Surv. Rep. 347: 2.

Irwin, M.E., L. Lyneborg. 1981a. The genera of Nearctic Therevidae. Bull. Ill. Nat. Hist. Surv. 32: 193-277.

Irwin, M.E., L. Lyneborg. 1981b. Therevidae. In "Manual of Nearctic Diptera" (J.F. McAlpine, B.V. Peterson, G.E. Shewell, H.J. Teskey, J.R. Vockeroth, D.M. Wood, Eds.), pp. 513-523. Research Branch, Agriculture Canada Monograph 1.b

Irwin, M.E., L. Lyneborg. 1989. Family Therevidae. In "Catalog of Diptera of the Australasian and Oceanian Regions" (N.L. Evenhuis, Ed.), pp. 353-358. Bishop Museum Special Publication 86. Bishop Museum Press, Honolulu.

Irwin, M.E., D.K. Yeates. 1998. "The Internal Hierarchy of the Therevidae and Biogeographic Patterns of Extant Higher Rank Taxa (Brachycera: Asiloidea)," Abstracts Volume, Fourth International Congress of Dipterology, p. 90. Oxford, United Kingdom.

Khambamphati, S. 1995. A phylogeny of cockroaches and related insects based on DNA sequence of mitochondrial ribosomal RNA genes. Proc. Natl. Acad. Sci. USA 92: 2017-2020.

Kitamoto, J., K. Sakamoto, K. Ozaki, Y. Mishina, 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.

Kocher, T.D., W.K. Thomas, A. Meyer, S.V. Edwards, S. Pääbo, F.X. Villablanca, A.C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86: 6196-6200.

Kojima, S., T. Hashimoto, M. Hasegawa, S. Murata, S. Ohta, H. Seki, N. Okada. 1993. Close phylogenetic relationship between Vestimentifera (tube worms) and Annelida revealed by the amino acid sequence of elongation factor-1-alpha. J. Mol. Evol. 37: 66-70.

Lopez, J.V., M. Culver, J.C. Stephens, W.E. Johnson, S.J. O'Brien. 1997. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. Mol. Biol. Evol. 14: 277-286.

Lyneborg, L. 1976. A revision of the therevine stiletto-flies (Diptera: Therevidae) of the Ethiopian region. Bull. Br. Mus. (Nat. Hist.) Entomol. 33: 191-346.

Mardulyn, P., S.A. Cameron. 1999. The major opsin in bees (Insecta: Hymenoptera): a promising nuclear gene for higher level phylogenetics. Mol. Phylogenet. Evol. 12: 168-176.

Martynov, V.I., M.B. Kostina, M. Yu. Feigina, A.I. Miroshnikov. 1983. A study of the molecular organization of visual rhodopsin in photoreceptor membranes by limited proteinolysis. Bioorg. Khim. 9: 734-745.

Meinertzhagen, I.A., R. Menzel, G. Kahle. 1983. The identification of spectral receptor types in the retina and lamina of the dragonfly *Sympetrum rubicundulum*. J. Comp. Physiol. 151: 295-310.

Metz, M.A. 2002. Systematics within the Therevinae (Insecta: Diptera: Therevidae). Ph.D. Dissertation. Department of Entomology, University of Illinois at Urbana-Champaign.

Montell, C., K. Jones, C. Zuker, G. Rubin. 1987. A second opsin gene expressed in the ultraviolet-sensitive R7 photoreceptor cells of *Drosophila melanogaster*. J. Neurosci. 7: 1558-1566.

Mullen, E., M. Akhtar. 1981. Topographic and active sites studies on bovine rhodopsin. FEBS Lett. 132: 261-264.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, H. Erlich. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor Symposium in Quantitative Biology 51: 263-73.

Mullis, K., F. Faloona. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. 155: 335-350.

Nagatomi, A. 1977. Classification of lower Brachycera (Diptera). J. Nat. Hist. 11: 321-335.

Nathans, J., D.S. Hogness. 1983. Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. Cell 34: 807-814.

Nathans, J., D.S. Hogness. 1984. Isolation and nucleotide sequence of the gene encoding human rhodopsin. Proc. Natl. Acad. Sci. U.S.A. 81: 4851-4855.

Nathans, J., D. Thomas, D.S. Hogness. 1986. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. Science 232: 193-202.

Naylor, G.J.P., W.M. Brown. 1998. *Amphioxus* mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. Syst. Biol. 47: 61-76.

O'Tousa, J., W. Baehr, R.L. Martin, J. Hirsch, W.L. Pak, M.L. Applebury. 1985. The *Drosophila ninaE* gene encodes an opsin. Cell 40: 839-850.

Ovchinnikov, Yu. A. 1982. Rhodopsin and bacteriorhodopsin: structure-function relationships. FEBS Lett. 148: 179-191.

Ovchinnikov, Yu. A., N.G. Abdulaev, M. Yu. Feigina, I.D. Artamonov, A.S. Bogachuk, A.S. Zolotarev, E.R. Eganyan, P.V. Kostetskii. 1983. Visual rhodopsin III: complete amino acid sequence and topography in the membrane. Bioorg. Khim. 9: 1331-1340.

Pashley, D.P., L.D. Ke. 1992. Sequence evolution in mitochondrial ribosomal and ND-1 genes in Lepidoptera: implications for phylogenetic analyses. Mol. Biol. Evol. 9: 1061-1075.

Paul, R., A. Steiner, R. Gemperlein. 1986. Spectral sensitivity of *Calliphora erythrocephala* and other insect species studied with Fourier interferometric stimulation (FIS). J. Comp. Physiol. A 158: 669-680.

Peitsch, D., A. Feitz, H. Hertl, J. de Souza, D.F. Ventura, R. Menzel. 1992. The spectral input systems of hymenopteran insects and their receptor-based colour vision. J. Comp. Physiol. A 170: 23-40.

Popp, M. P., R. Grisshammer, P.A. Hargrave, W. Smith. 1996. Ant opsins: Sequences from the Saharan silver ant and the carpenter ant. Invertebr. Neurosci. 1: 323-329.

Reiger, J.C., J.W. Schultz. 1997. Molecular phylogeny of the major arthropod groups indicates polyphyly of the crustaceans and a new hypothesis for the origin of the hexapods. Mol. Biol. Evol. 14: 902-913.

Rokas, A., J.A.A. Nylander, F. Ronquist, G.N. Stone. 2002. A maximum-likelihood analysis of eight phylogenetic markers in gallwasps (Hymenoptera: Cynipidae): implications for insect phylogenetic studies. Mol. Phylogenet. Evol. 22: 206-219.

Saiki, R., S. Scharf, F. Faloona, K. Mullis, G. Horn, H. Erlich. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230: 1350-1354.

Seki, T., K. Vogt. 1998. Evolutionary aspects of the diversity of visual pigment chromophores in the class Insecta. Comp. Biochem. Physiol. B 119: 53-64.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87: 651-701.

Smith, S.W., R. Overbeek, C.R. Woese, W. Gilbert, P.M. Gillevet. 1994. The genetic data environment and expandable GUI for multiple sequence analysis. Comput. Appl. Biosci. 10: 671-675.

Swofford, D.L., G.J. Olsen, P.J. Waddell, D.M. Hillis. 1996. Phylogenetic Inference. In "Molecular Systematics" (D.M. Hillis, C. Moritz, B.K. Mable, Eds.), pp. 407-514. Sinauer Associates, Sunderland.

Swofford, D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer, Sunderland, MA.

Towner, P., W. Gärtner. 1994. The primary structure of mantid opsin. Gene 143: 227-231.

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

Wald, G. 1968. The molecular basis of visual excitation. Nature 219: 800-807.

Wiegmann, B.M., C. Mitter, J.C. Reiger, T.P. Friedlander, D.M. Wagner, E.S. Nielsen. 2000. Nuclear genes resolve Mesozoic-aged divergences in the insect order Lepidoptera. Mol. Phylogenet. Evol. 15: 242-259.

Wiegmann, B.M., D.K. Yeates, J.L. Thorne, H. Kishino. Time flies: A new molecular timescale for brachyceran fly evolution without a clock. Syst. Biol., in press.

Winterton, S.L., M.E. Irwin. 1999. *Laxotela*- a new genus of Therevidae (Diptera) from Australia. Entomol. Scand. 30: 299-310.

Winterton, S.L., M.E. Irwin, D.K. Yeates. 1999. Systematics of *Nanexila* Winterton & Irwin, gen. nov. (Diptera: Therevidae) from Australia. Invertebr. Taxon. 13: 237-308.

Winterton, S.L., J.H. Skevington, M.E. Irwin, D.K. Yeates. 2000. Phylogenetic revision of *Bonjeania* Irwin & Lyneborg (Diptera: Therevidae). Syst. Entomol. 25: 295-324.

Winterton, S.L., L. Yang, B.M. Wiegmann, D.K. Yeates. 2001. Phylogenetic revision of Agapophytinae subf.n. (Diptera: Therevidae) based on molecular and morphological evidence. Syst. Entomol. 26: 173-211.

Woodley, N.E. 1989. Phylogeny and classification of the "Orthorrhaphous" Brachycera. In "Manual of the Nearctic Diptera, Vol. 3" (J.F. McAlpine, Ed.), pp. 1371-1395. Monograph No. 32, Research Branch, Agriculture Canada, Ottawa.

Yang, E-C, D. Osorio. 1991. Spectral sensitivities of photoreceptors and lamina monopolar cells in the dragonfly, *Hemicordulia tau*. J. Comp. Physiol. A 169: 663-669.

Yang, L.L. 2000. Molecular phylogenetics of the Therevidae and their position among the families of the Asiloidea (Insecta: Diptera). Ph.D. Thesis. Department of Entomology, North Carolina State University.

Yang, L.L., B.M. Wiegmann, D.K. Yeates, M.E. Irwin. 2000. Higher-level phylogeny of the Therevidae (Diptera: Insecta) based on 28S ribosomal and elongation factor-1α gene sequences. Mol. Phylogenet. Evol. 15: 440-451.

Yeates, D.K. 1994. The cladistics and classification of the Bombyliidae (Diptera: Asiloidea). Bull. Am. Mus. Nat. Hist. 219: 1-191.

Yeates, D.K. 2002. Relationships of extant lower Brachycera (Diptera): A quantitative synthesis of morphological characters. Zool. Scr. 31: 105-121.

Yeates, D.K., M.E. Irwin, B.M. Wiegmann. Ocoidae, a new family of asiloid flies (Diptera: Brachycera: Asiloidea), based on *Ocoa chilensis* gen. and sp. nov. from Chile, South America. Syst. Ent., submitted.

Yeates, D.K., B.M. Wiegmann. 1999. Congruence and controversy: Toward a higher-level phylogeny of Diptera. Annu. Rev. Entomol. 44: 397-428.

Yokoyama, S. 2000. Color vision of the coelacanth (*Latimeria: chalumnae*) and adaptive evolution of rhodopsin (RH1) and rhodopsin-like (RH2) pigments. J. Hered. 91: 215-220.

Yokoyama, R., S. Yokoyama. 1996. Adaptive evolution of photoreceptors and visual pigments in vertebrates. Annu. Rev. Ecol. Syst. 27: 543-567.

Zardoya, R., A. Meyer. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol. Biol. Evol. 13: 933-942.

Zuker, C.S. 1996. The biology of vision in *Drosophila*. Proc. Natl. Acad. Sci. 93: 571-576.

Zuker, C., A.F. Cowman, G.M. Rubin. 1985. Isolation and structure of a rhodopsin gene from *D. melanogaster*. Cell 40: 851-858.

Zuker, C.S., C. Montell, K. Jones, T. Laverty, G.M. Rubin. 1987. A rhodopsin gene expressed in photoreceptor cell R7 of the *Drosophila* eye: homologies with other signal-transducing molecules. J. Neurosci. 7: 1550-1557.

Zuckerkandl, E., L. Pauling. 1962. Molecular disease, evolution and genic heterogeneity. In "Horizons in biochemistry" (M. Kash, B. Pullman, Eds.), pp. 189-225. Academic Press, New York.

Table 1. Taxa sampled for opsin from Bombyliidae, Apsilocephalidae, Scenopinidae, Empididae, and Therevidae.

Taxon	GenBank Accession Number	MEI Number	Geographical Source
EMPIDOID OUTGROUPS			
Empididae			
Empis	AY267594	N/A	North Carolina
Bicellaria	AY267592	N/A	North Carolina
ASILOID OUTGROUPS			
Bombyliidae			
Mythicomyia	AY267578	N/A	California
Apsiliocephalidae			
Apsilocephala	AY267591	103270	California
Clesthentia	AY267593	N/A	Tasmania
Scenopinidae			
Stenomphrale teutankhameni	AY267587	029578	Israel
Alloxytropus	AY267590	050694	Israel
INGROUP: Therevidae			
Phycinae			
Hemigephyra	AY267573	124388	S. Africa
Henicomyia	AY267574	139545	Arizona
Ruppellia multisetosis	AY267585	011183	Madagascar
Phycus	AY267583	050693	Israel
Neotabuda	AY267579	089473	S. Africa
Ataenogera abdominalis	AY267563	124662	Argentina
Ataenogera sp.	AY267562	N/A	
Pherocera	AY267584	N/A	California
Agapophytinae (sensu lato)			
Agapophytus bicolor	AY267559	N/A	Australia
Agapophytus adonis	AY267558	N/A	Australia
Acraspisa sp. 1	AY267555	089456	Australia
Acraspisa sp. 2	AY267556	N/A	Australia
Acraspisa sp. 3	AY267554	089470	Australia
Actenomeros	AY267557	N/A	Australia
Patanothrix skevingtoni	AY267567	053724	Australia
Patanothrix wilsoni	AY267569	N/A	Australia
Entesia	AY267572	106357	Chile
Melanothereva	AY267577	106857	Chile
New Genus M	AY267566	N/A	Australia
New Genus S	AY267586	N/A	Australia
Therevinae			
Anabarhynchus sp. 2	AY267561	N/A	New Caledonia
Anabarhynchus sp. 1	AY267560	N/A	Australia
Brachylinga sp.	AY267564	090244	Guatemala
Brachylinga tepocacae	AY267565	090257 / 090274	Guatemala
Ozodiceromyia flavipennis	AY267580	089437	Illinois
Ozodiceromyia metallica	AY267581	089520	Colorado
Cyclotelus pictipennis	AY267571	089439	Illinois
Cyclotelus sp.	AY267570	N/A	

Table 1. (Continued)

Taxon	GenBank Accession Number	MEI Number	Geographical Source
Therevinae			
Thereva sp.	AY267589	089516	Colorado
Thereva mirabilis	AY267588	119946	N. Africa
Irwiniella velutina	AY267576	011319	Madagascar
Penniverpa	AY267582	090271	Guatemala
Chromolepida pruinosa	AY267568	090266 / 090267	Guatemala
Hoplosathe frauenfeldi	AY267575 ("ops" copy) AY267595 ("lw" copy)	050690	Israel

Table 2. Sequences from different opsin copies of insects retrieved from GenBank.

Insect order	Species	Opsin "Copy"	Accession Number	Reference
Diptera	*Drosophila melanogaster*	Rh1	K02315	O'Tousa *et al*., 1985
	D. melanogaster	Rh2	M12896	Cowman *et al*., 1986
	D. melanogaster	Rh3	M17718	Zuker *et al.*, 1987
	D. melanogaster	Rh4	M17719	Montell *et al*., 1987
	D. melanogaster	Rh5	U67905	Chou *et al*., 1996
	D. pseudoobscura	Rh1	X65877	Carulli and Hartl, 1992
	D. pseudoobscura	Rh2	X65878	Carulli and Hartl, 1992
	D. pseudoobscura	Rh3	X65879	Carulli and Hartl, 1992
	D. pseudoobscura	Rh4	X65880	Carulli and Hartl, 1992
	Calliphora vicina	Ops1	M58334	Huber *et al.*, 1990
	Anopheles gambiae	Ops	Y17705	Arca *et al*., 1999
Hymenoptera	*Apis mellifera*	Green wl	U26026	Chang *et al.*,1996
	A. mellifera	UV	AF004169	Townson *et al*.,1998
	A. mellifera	Blue wl	AF004168	Townson *et al*.,1998
	Camponotus abdominalis	Ops1	U32502	Popp *et al.*, 1996
	C. abdominalis	Ops2	AF042788	Smith *et al.*, unpublished
	Cataglyphis bombycinus	Ops1	U32501	Popp *et al.*, 1996
	C. bombycinus	Ops2	AF042787	Smith *et al.*, unpublished
Lepidoptera	*Manduca sexta*	Ops1	L78080	Chase *et al.*, 1997
	M. sexta	Ops2	L78081	Chase *et al.*, 1997
	M. sexta	Ops3	AD001674	Chase *et al.*, 1997
	Papilio xuthus	Rh1	AB007423	Kitamoto *et al.*, 1998
	P. xuthus	Rh2	AB007424	Kitamoto *et al.*, 1998
	P. xuthus	Rh3	AB007425	Kitamoto *et al.*, 1998
	P. xuthus	Rh4	AB028217	Kitamoto *et al.*, unpublished
	P. xuthus	Rh5	AB02818	Kitamoto *et al.*, unpublished
Orthoptera	*Schistocerca gregaria*	Ops1	X80072	Towner *et al.*, unpublished
	S. gregaria	Ops2	X80071	Towner *et al.*, unpublished
Mantodea	*Sphodromantis* sp.	Ops	X71665	Towner and Gartner, 1994
Hemiptera	*Megoura viciae*	UV	AF189715	Gao *et al*., 2000
	M. viciae	Long wl	AF189714	Gao *et al*., 2000

Table 3. Primer sequences of opsin used for PCR and sequencing. Degenerate positions are noted by their IUB single-letter code: R=A/G; Y=C/T; M=A/C; N= A/G/C/T.

Gene	Primer	Sequence (5' → 3')
Opsin	ops102f	ATGATHACNAAYACNCCNATGAT
	ops336r	CTDATNCCRTANACDATNGGRTT
	ops1f	ATTTGGTCGATGTGCATGAT
	ops3r	GCAATTATGAACCAATAAG
	lwrhf	AATTGCTATTAYGARACNTGGGT
	M13(-21)-lwrhf	TGTAAAACGACGGCCAGTAATTGCTATTAYGARACNTGGGT
	M13f	TGTAAAACGACGGCCAGT
	lwrhr	ATATGGAGTCCANGCCATRAACCA
	M13-lwrhr	CAGGAAACAGCTATGACCATATGGAGTCCANGCCATRAACCA
	M13r	CAGGAAACAGCTATGACC

Table 4. Tree statistics for unweighted maximum parsimony trees (MPTs).

Taxon Analysis Set	All Opsins 123	GenBank 123	GenBank 12	Diptera 123	Therevidae 123lw	Therevidae 123
MPT length	3240	2201	1027	1865	1045	718
No. MPTs	15	1	1	25	2	4
CI	0.241	0.348	0.462	0.362	0.365	0.373
RI	0.550	0.505	0.645	0.543	0.543	0.528

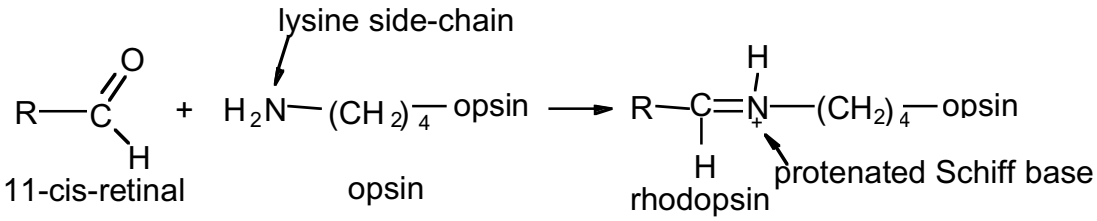

Figure 1. Common structural plan of visual pigments.

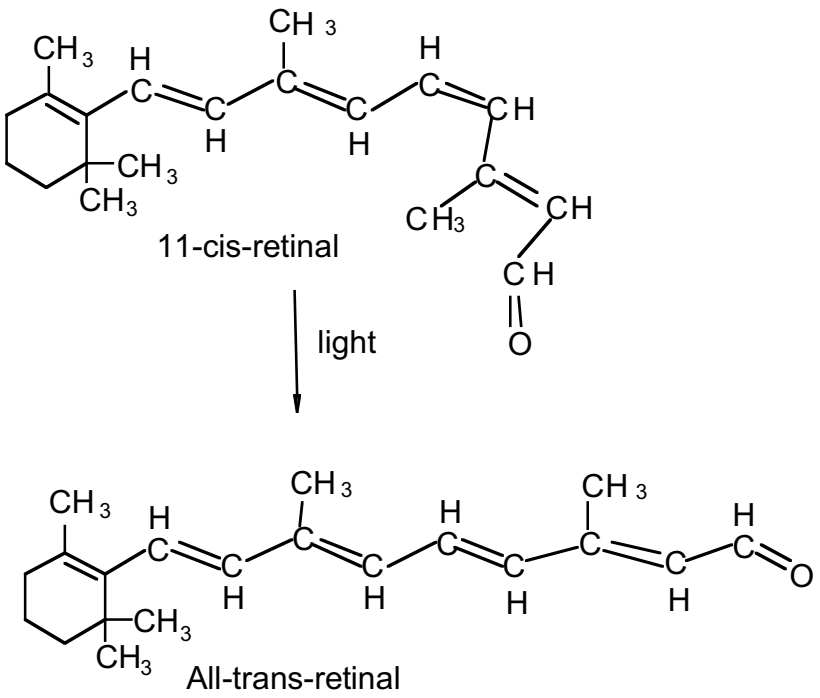

Figure 2. Activation of a visual pigment through the absorption of a photon of light.

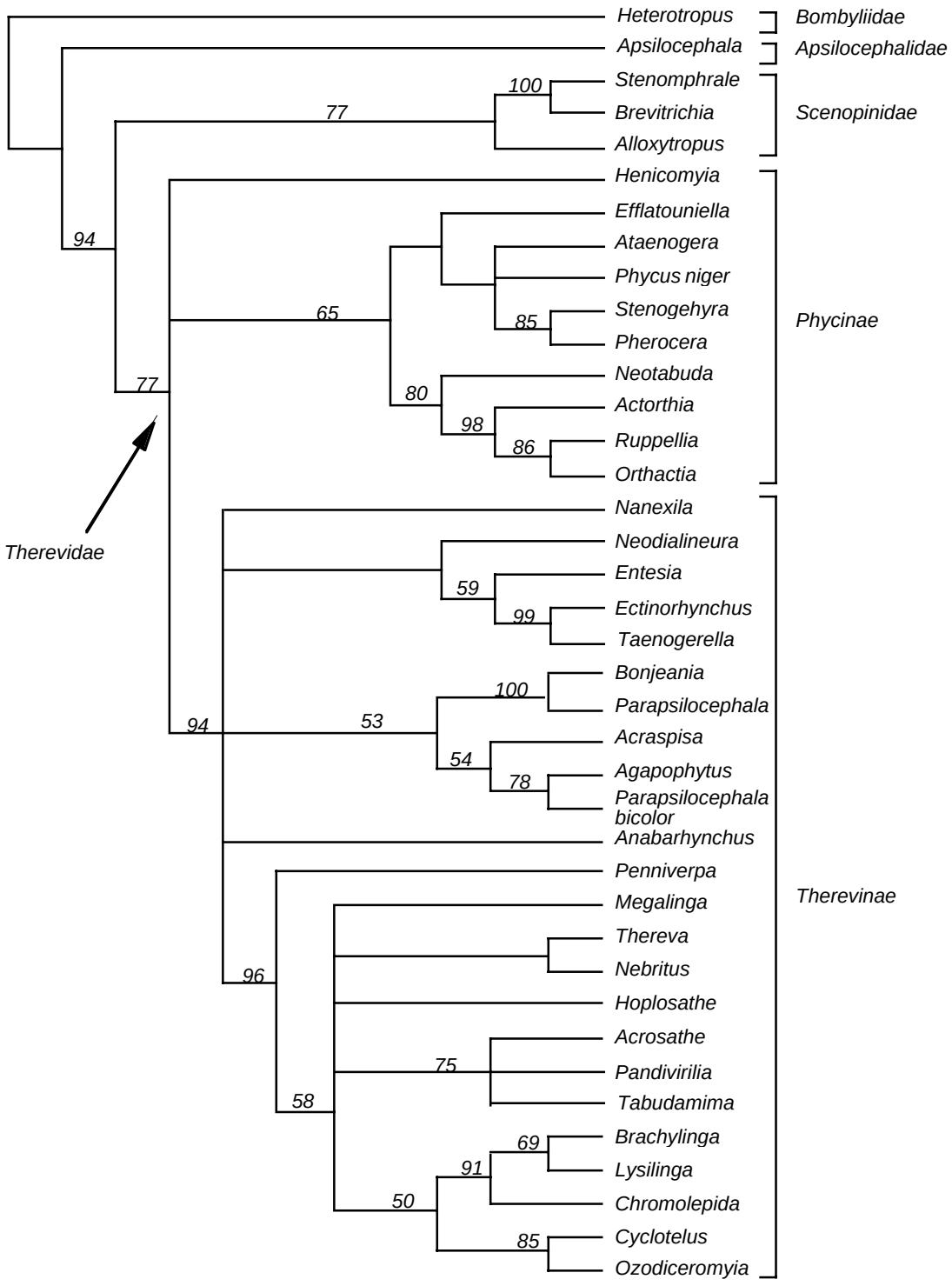

Figure 3. Phylogenetic relationships based on combined 28S rDNA and EF-1α gene sequences of 39 genera from study conducted by Yang et al. (2000). Strict consensus of 14 EPTs with the bootstrap values (200 replicates) shown above each branch.

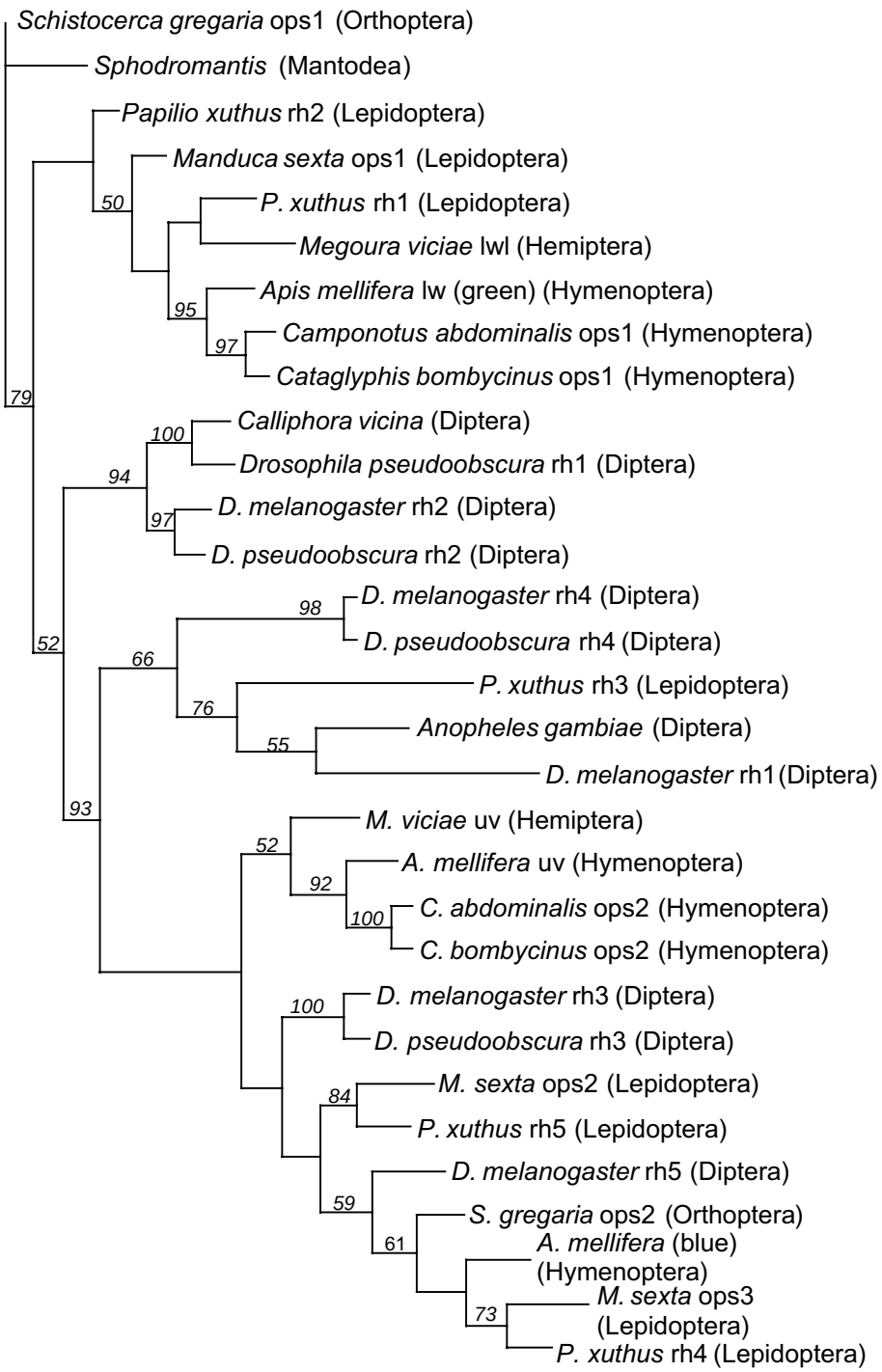

Figure 4. Single MPT demonstrating the phylogenetic relationships between published insect opsin sequences (outgroup = *Schistocerca gregaria*). The tree was estimated using unweighted parsimony analysis of nucleotide sequences, implemented in PAUP*, version 4.0. Tree length = 2201, CI = 0.348, RI = 0.505. Bootstrap percentages (calculated with 1000 replicate heuristic searches) ≥ 50% are shown above each branch.

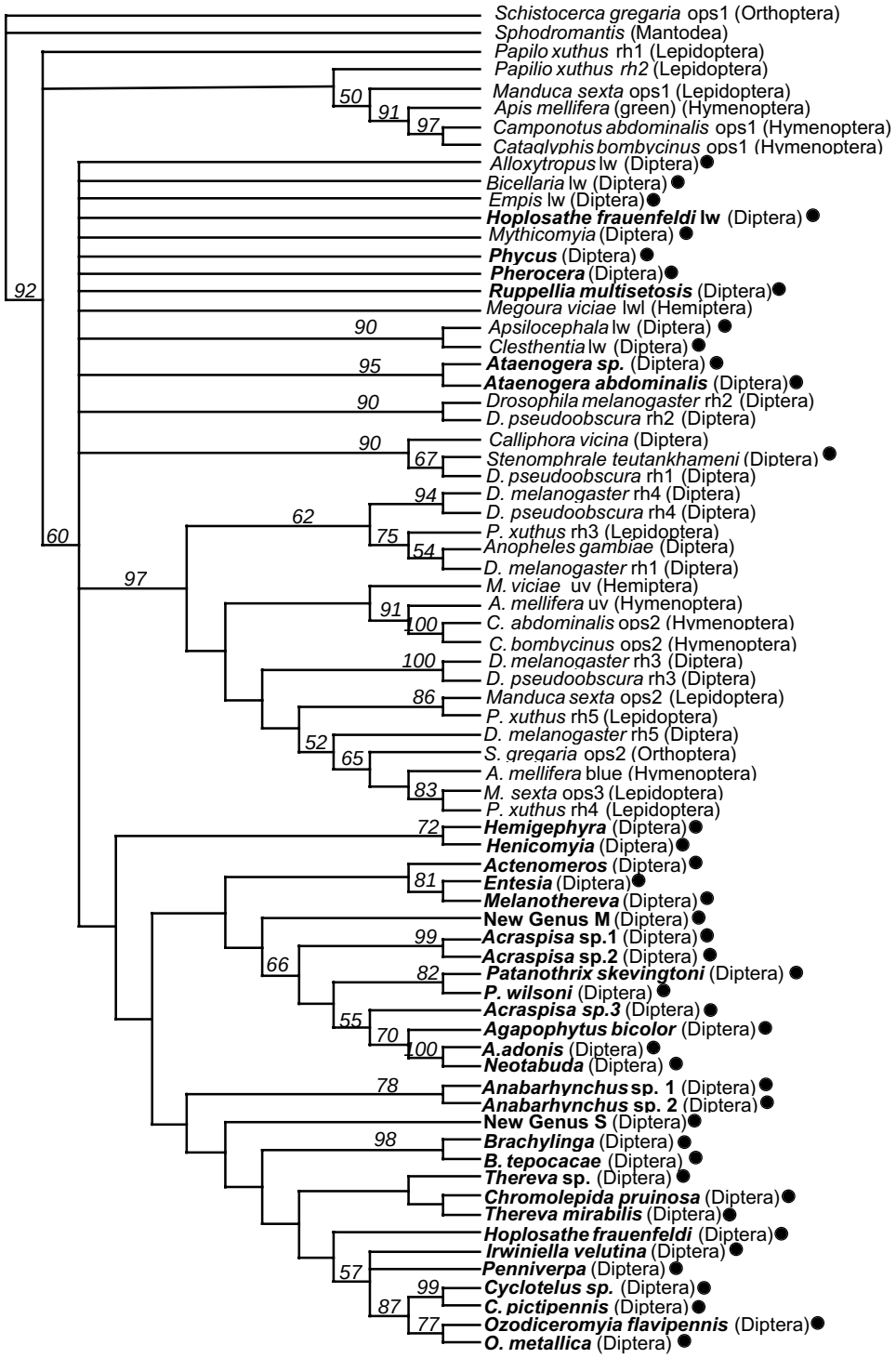

Figure 5. Strict consensus of 15 EPTs demonstrating phylogenetic relationships between published insect opsin sequences, including sequences obtained in this study, represented by closed circles, therevids in bold type (outgroup = *Schistocerca gregaria*). The tree was estimated using unweighted parsimony analysis of nucleotide sequences, implemented in PAUP*, version 4.0. Tree length = 3240, CI = 0.241, RI = 0.550. Bootstrap percentages (calculated with 1000 replicate heuristic searches) ≥ 50% are shown above each branch.

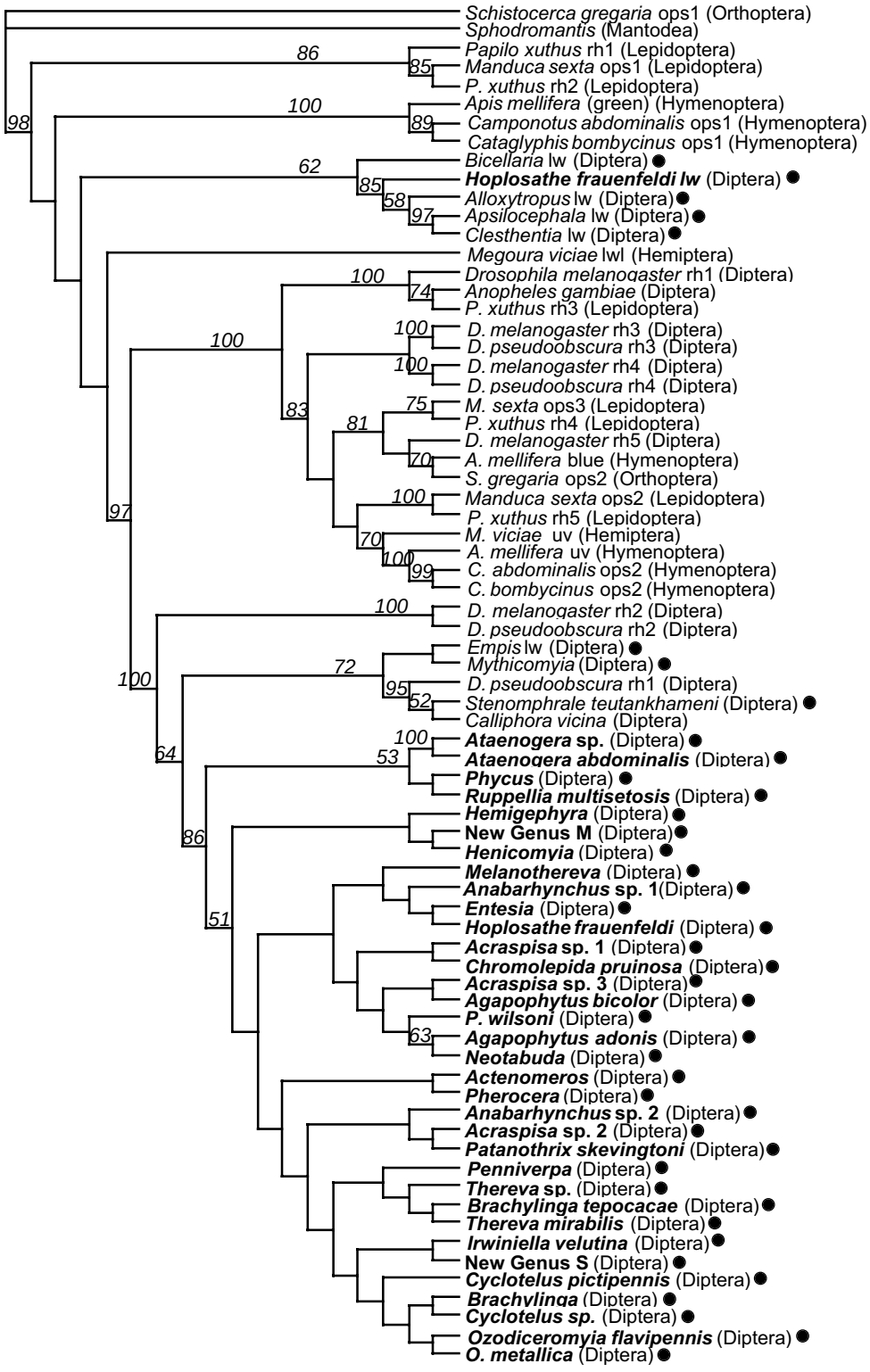

Figure 6. Neighbor-joining tree (excluding third positions) demonstrating phylogenetic relationships between published insect opsin sequences, including sequences obtained in this study represented by closed circles, therevids in bold type (outgroup = *Schistocerca gregaria*). This tree was estimated using minimum evolution distance analysis of nucleotide sequences in PAUP*, version 4.0. Bootstrap percentages (calculated with 100 replicates and TBR branch swapping) ≥ 50% are shown above each branch.

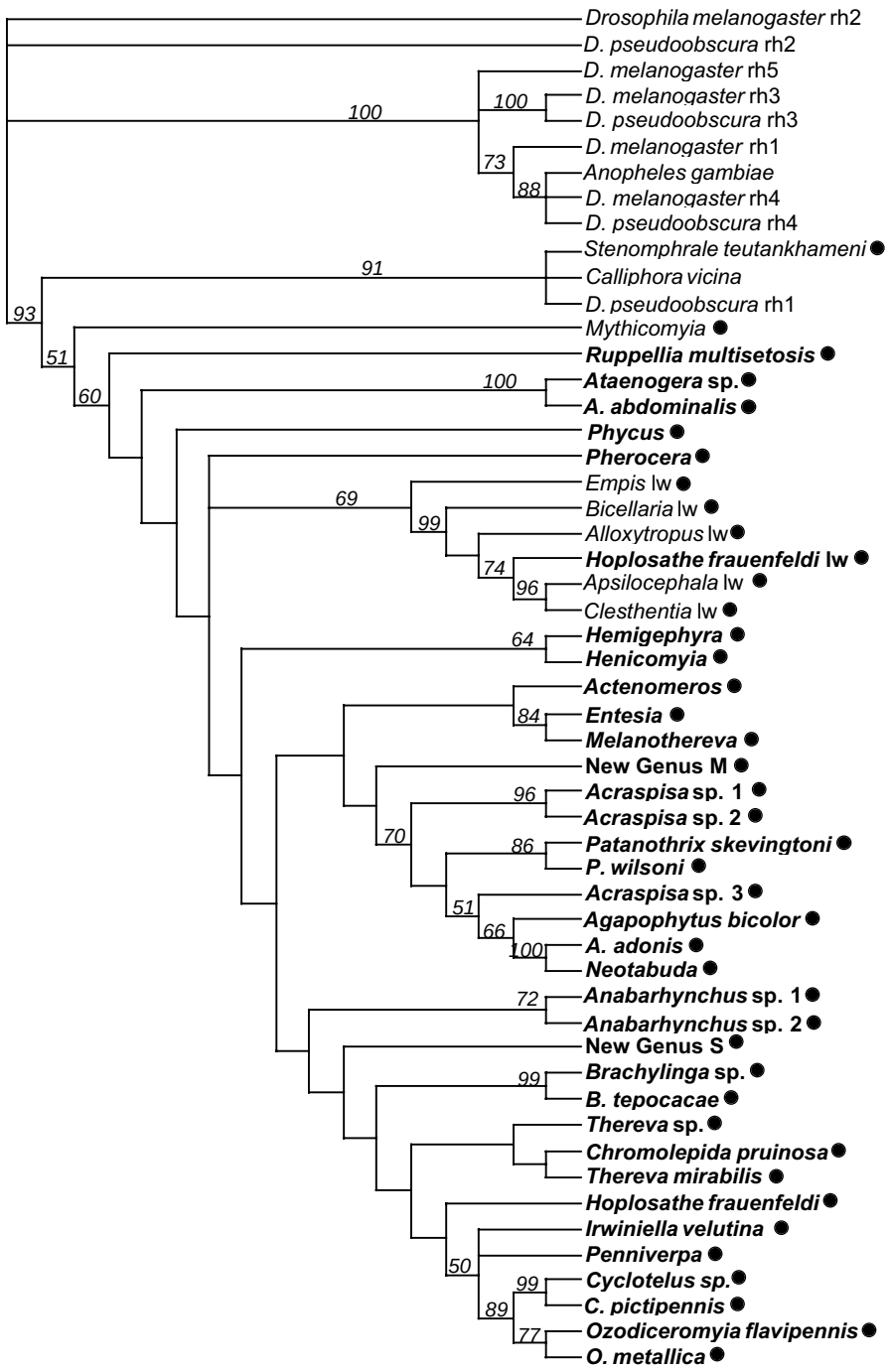

Figure 7. Strict consensus of 25 EPTs demonstrating phylogenetic relationships between different dipteran opsin sequences, including sequences obtained in this study, represented by closed circles, therevids in bold type (outgroup = GenBank sequences). The tree was estimated using unweighted parsimony analysis of nucleotide sequences, implemented in PAUP*, version 4.0. Tree length = 1865, CI = 0.362, RI = 0.543. Bootstrap percentages (calculated with 1000 replicate heuristic searches) ≥ 50% are shown above each branch.

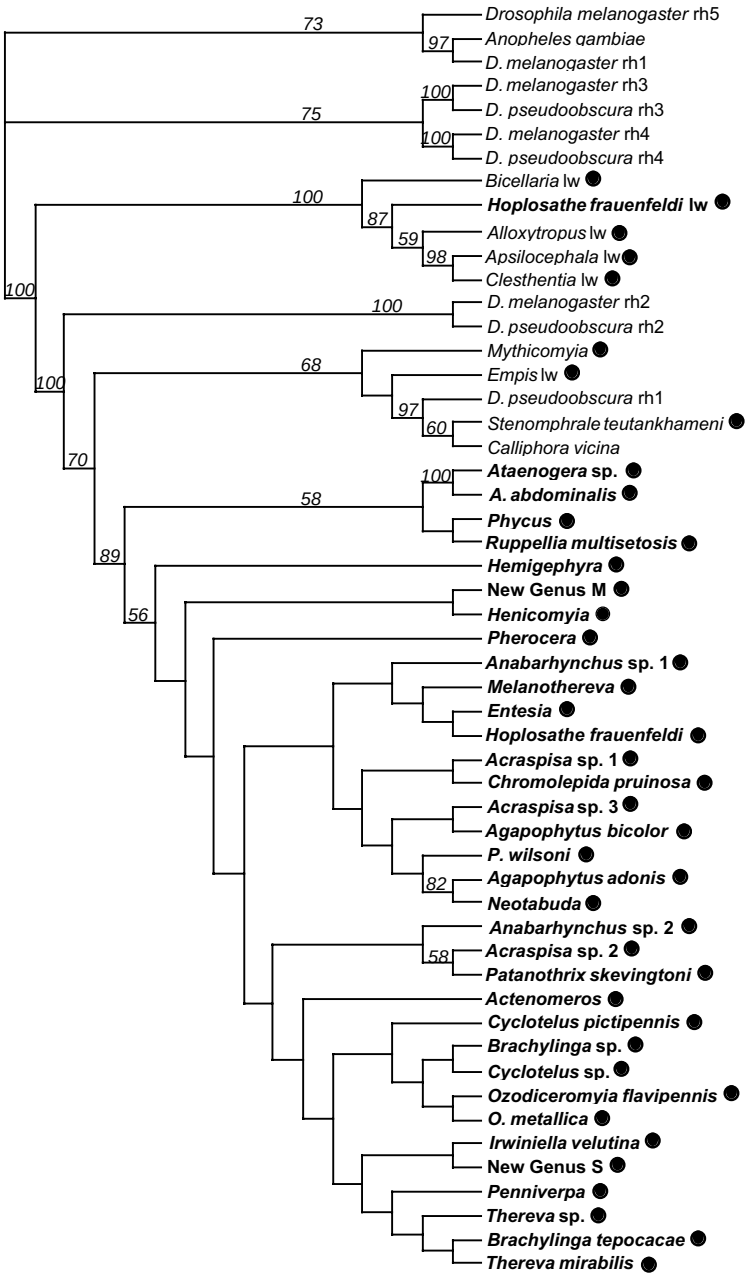

Figure 8. Neighbor-joining tree (excluding third positions) demonstrating phylogenetic relationships between different dipteran opsin sequences, including sequences obtained in this study represented by closed circles, therevids in bold type (outgroup = GenBank sequences). This tree was estimated using minimum evolution distance analysis of nucleotide sequences in PAUP*, version 4.0. Bootstrap percentages (calculated with 100 replicates and TBR branch swapping) ≥ 50% are shown above each branch.

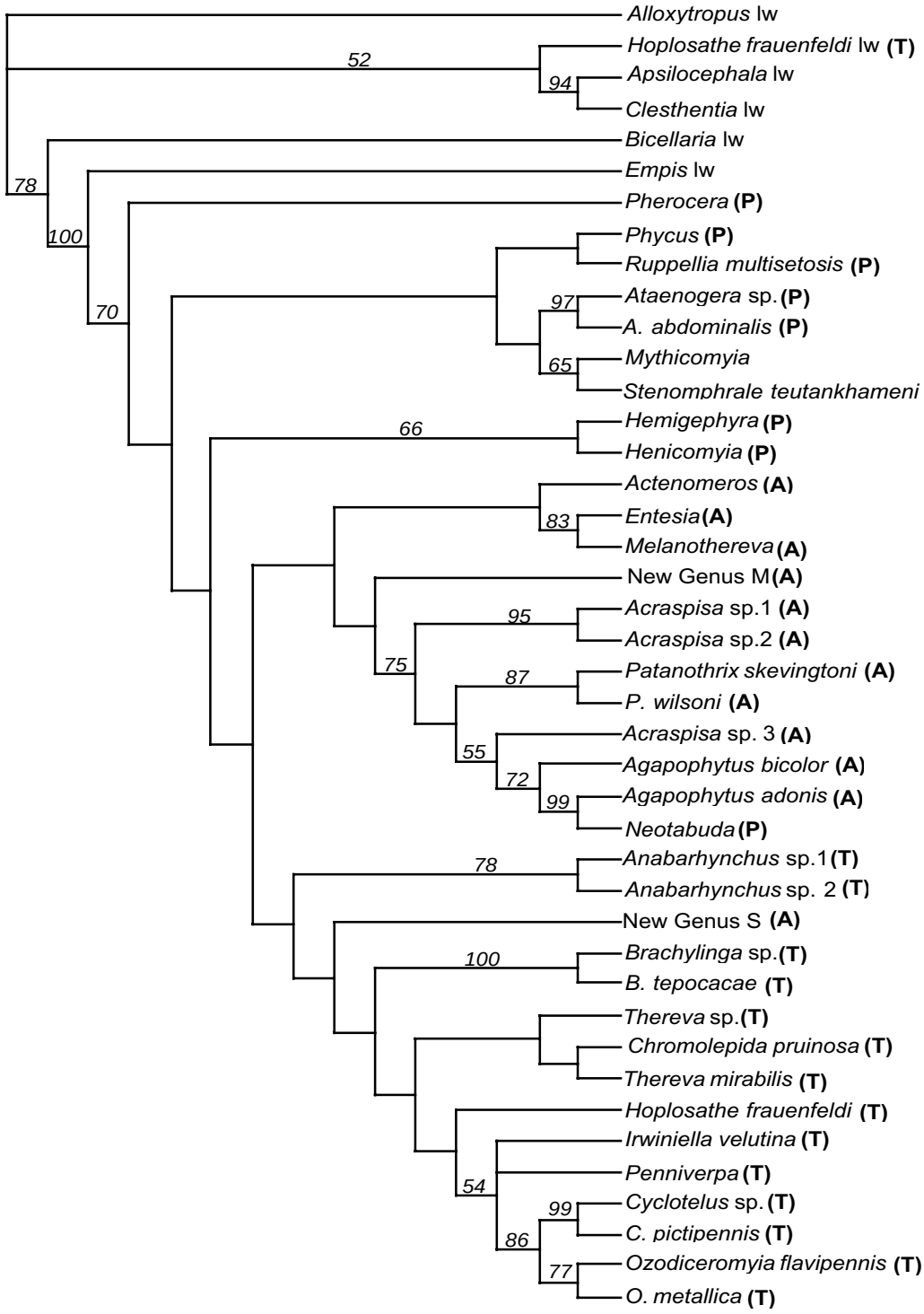

Figure 9. Strict consensus of two EPTs demonstrating the phylogenetic relationships between different opsin sequences of Therevidae and close relatives ("lw" copies included). Outgroups = *Alloxytropus* lw, *Apsilocephala* lw, *Bicellaria* lw, *Clesthentia* lw, *Empis* lw, *Mythicomyia*, and *Stenomphrale teutankhameni*. Subfamilies denoted as follows: Agapophytinae = A, Phycinae = P, and Therevinae = T. The tree was estimated using unweighted parsimony analysis of nucleotide sequences, implemented in PAUP*, version 4.0. Tree length = 1045, CI = 0.365, RI = 0.543. Bootstrap percentages (calculated with 1000 replicate heuristic searches) ≥ 50% are shown above each branch.

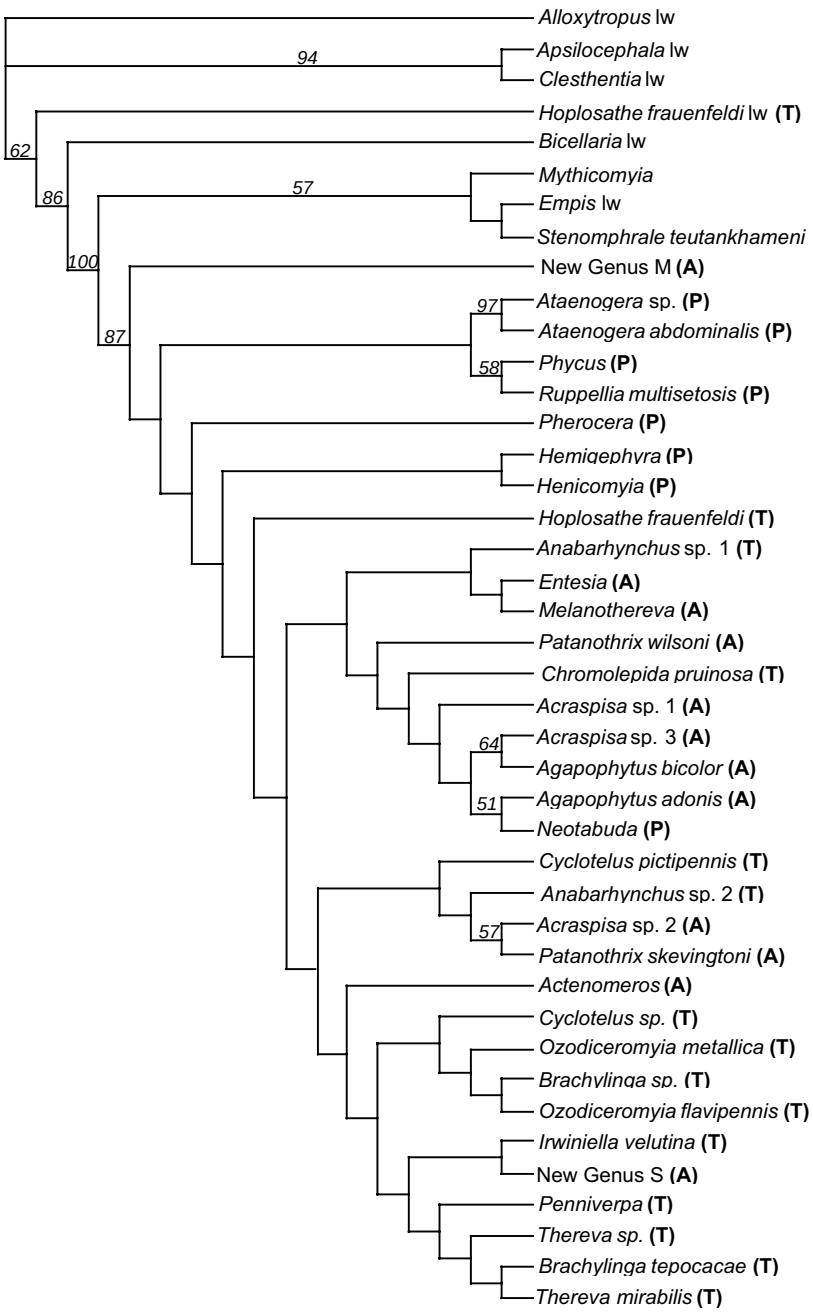

Figure 10. Neighbor-joining tree (excluding third positions) demonstrating phylogenetic relationships between different opsin sequences of Therevidae and close relatives ("lw" copies included). Outgroups = *Alloxytropus* lw, *Apsilocephala* lw, *Bicellaria* lw, *Clesthentia* lw, *Empis* lw, *Mythicomyia*, and *Stenomphrale teutankhameni*. Subfamilies denoted as follows: Agapophytinae = A, Phycinae = P, and Therevinae = T. This tree was estimated using minimum evolution distance analysis of nucleotide sequences in PAUP*, version 4.0. Bootstrap percentages (calculated with 100 replicates and TBR branch swapping) ≥ 50% are shown above each branch.

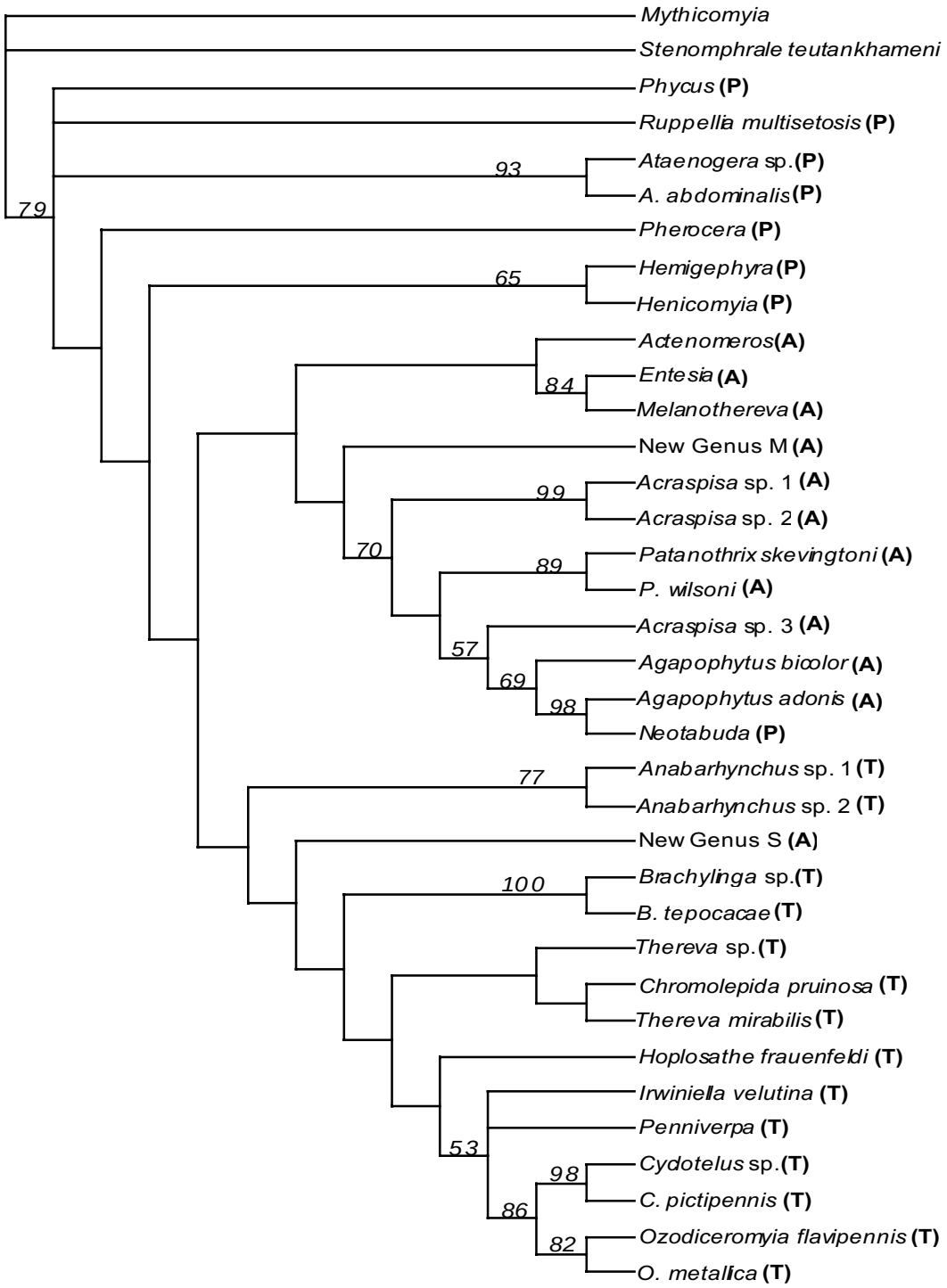

Figure 11. Strict consensus of four EPTs demonstrating the phylogenetic relationships between different opsin sequences of Therevidae and close relatives ("lw" copies excluded). Outgroups = *Mythicomyia* and *Stenomphrale teutankhameni*. Subfamilies denoted as follows: Agapophytinae = A, Phycinae = P, and Therevinae = T. The tree was estimated using unweighted parsimony analysis of nucleotide sequences, implemented in PAUP*, version 4.0. Tree length = 718, CI = 0.373, RI = 0.528. Bootstrap percentages (calculated with 1000 replicate heuristic searches) ≥ 50% are shown above each branch.

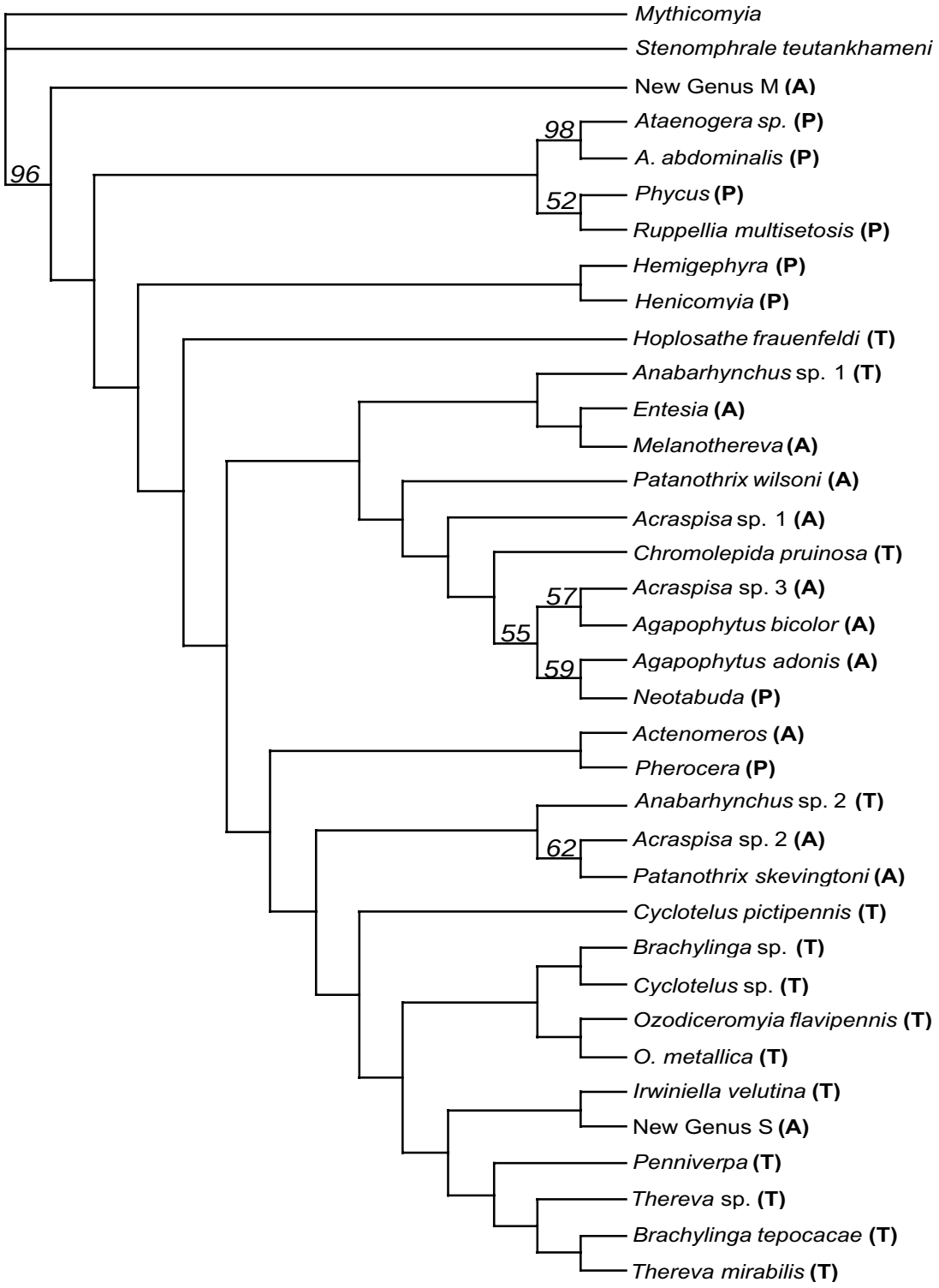

Figure 12. Neighbor-joining tree (excluding third positions) demonstrating phylogenetic relationships between different opsin sequences of Therevidae and close relatives ("lw" copies excluded). Outgroups = *Mythicomyia*, and *Stenomphrale teutankhameni*. Subfamilies denoted as follows: Agapophytinae = A, Phycinae = P, and Therevinae = T. This tree was estimated using minimum evolution distance analysis of nucleotide sequences in PAUP*, version 4.0. Bootstrap percentages (calculated with 100 replicates and TBR branch swapping) ≥ 50% are shown above each branch.

PHYLOGENETIC UTILITY OF TWO NEW NUCLEAR GENES, OPSIN AND CAD, WITHIN THE STILETTO FLIES (DIPTERA: THEREVIDAE)

CHAPTER 2

Therevid Phylogeny Based on Multiple Nuclear Genes

ABSTRACT

In 2000, the first molecular phylogeny of the Therevidae (Diptera: Asiloidea) was published based on nucleotide sequences of elongation factor-1α (EF-1α) and 28S ribosomal DNA (28S rDNA). That study by Yang, Wiegmann, Yeates, and Irwin produced a well-supported classification, though they had insufficient molecular data to conclusively specify relationships within and among lineages of the Therevinae. A need for multiple phylogenetic markers to reconstruct evolutionary relationships has become increasingly apparent both within the Therevidae and insect systematics as a whole. In this study we applied two new nuclear, protein-encoding genes, opsin and CAD, in combination with EF-1α and 28S rDNA, to reconstruct evolutionary relationships among the major lineages of Therevidae for which previous molecular evidence has been insufficient.

Our combined analyses of these nucleotide sequences provide phylogenetic resolution that is highly concordant with the current therevid classification. We also demonstrate that the phycine tribe Xestomyzini is monophyletic and could represent a major therevid clade. The new molecular dataset generated from this study can be used to build a larger dataset and can also be used in conjunction with a morphological dataset that is currently being generated within the therevid PEET project

INTRODUCTION

Insect molecular systematists recognize that multiple phylogenetic markers are needed to elucidate relationships among major insect groups, especially markers informative for the great radiations of the Mesozoic (~65-250mya). Evolutionary relationships are best evaluated by using datasets from distinct sources and from character sets generated under different selective pressures (Wheeler et al. 1993, Dowton & Austin 2002). The majority of recent insect systematics studies incorporate various combinations of markers from multiple genomes (Flook et al. 1999, Wiegmann et al. 2000a, Yang et al. 2000, Abraham et al. 2001, Baker et al. 2001, Caterino et al. 2001, Krzywinski et al. 2001, Buckley et al. 2002, Johnson & Whiting 2002, Rokas et al. 2002), morphology (Whiting et al. 1997, Cameron & Mardulyn 2001, Joy & Conn 2001, Kjer et al. 2001, Wiegmann et al. 2001) and biogeography (Weekers et al. 2001).

Yang et al. (2000) were the first workers to produce a molecular phylogeny of the Therevidae (Diptera: Asiloidea) from nucleotide sequences of elongation factor-1α (EF-1α) and 28S ribosomal DNA (28S rDNA) (Fig. 1). Yang et al. (2000) produced a well-supported classification that recovered a monophyletic Therevidae comprised of two major clades corresponding with Irwin & Lyneborg's (1981a) division of the family into Phycinae and Therevinae. Nevertheless, Yang et al. (2000) had insufficient molecular evidence to conclusively specify relationships within and among lineages of the Therevinae.

In this study we apply two new nuclear, protein-encoding genes, opsin and CAD, in combination with EF-1α and 28S rDNA, to reconstruct evolutionary relationships among the major lineages of Therevidae. Until recently, previous studies of opsin as a phylogenetic marker in insects have been limited to relationships within the Hymenoptera (Mardulyn & Cameron 1999, Cameron & Mardulyn 2001, Ascher et al. 2001, Rokas et al. 2002). Here, we also present the first application of the nuclear gene CAD to phylogenetic analysis of the dipteran family Therevidae. Previously, phylogenetic work with CAD has been limited to investigations of the carbomoylphosphate synthetase (CPS) exon within the CAD locus across deep divergences

among kingdoms and within microbes (van den Hoff et al. 1995, Lawson et al. 1996). More recently, several regions of CAD, including the CPSase domain, have been used to successfully recover relationships among empidoid and cyclorrhaphan Diptera (Moulton and Wiegmann, unpublished results).

Phylogenetic Markers

Opsin is a multiple copy, nuclear visual pigment gene that encodes for light absorbing proteins, which mediate color vision. A sizeable amount is known about opsins and their role in color vision, because the system has been studied extensively at both the physiological and biochemical levels (Ovchinnikov 1982, Hargrave et al. 1983, Martynov et al. 1983, Mullen & Akhtar 1983, Nathans & Hogness 1983, Ovchinnikov et al. 1983, Barclay & Findlay 1984, Nathans & Hogness 1984, O'Tousa et al. 1985, Zuker et al. 1985). A visual pigment is composed of a light sensitive retinal-based chromophore (11-*cis* retinal) and an opsin protein (Briscoe & Chittka 2001). Opsin genes code for membrane embedded receptors that bind light sensitive properties through a single retinal chromophore covalently attached to a lysine residue in the opsin through a Schiff base linkage (Chang et al. 1995).

The combination of the number of visual pigments and their range of spectral absorption determine an organism's visual sensitivity spectrum (Crandall & Cronin 1997). The visual sensitivity spectrum is likely to be largely influenced by the environment (Crandall & Cronin 1997). The number of known visual pigments in organisms' ranges from one to ten and the spectral absorption (λ_{max}) ranges from 350 to 620 nm (Chang et al. 1995). These diverse ranges in number of visual pigments and spectral absorption are supported by an underlying genetic variation in the visual pigment genes (Chang et al. 1995).

Opsin (LW *Rh*) is approximately 1597 base pairs (bp) long in *Drosophila melanogaster* (Zuker et al. 1985) and has been reported to be phylogenetically informative at divergence times between 50 and 150 mya (Mardulyn & Cameron 1999, Cameron & Mardulyn 2001, Ascher et al. 2001, Rokas et al. 2002).

CAD, also known as the *rudimentary* gene in *Drosophila melanogaster,* is a single copy, multienzymatic protein that encodes the first three enzymatic activities in the pyrimidine biosynthesis pathway (Rawls & Porter 1979, Simmons et al. 1999). These catalytic enzymes are carbamyl phosphate synthetase (CPSase), aspartate transcarbamylase (ATCase), and dihydroorotase (DHOase), respectively (Nørby 1973, Jarry & Falk 1974, Rawls & Fristrom 1975, Freund et al. 1986; Fig. 2).

CAD is thought to be a potentially useful phylogenetic marker for the following reasons: CAD is a large (~6708 bp), single-copy gene; dipteran primers are available for the CPSase domain (J. K. Moulton, University of Tennessee (UTN); B. Wiegmann, North Carolina State University (NCSU)); and regions within CAD (CPSase domain) range from very conserved to quite variable and thus are potentially useful for both deeper and more recent divergences (Moulton pers. com.).

In addition to the above mentioned newly characterized markers, we also used 28S rDNA and EF-1α, both of which are well known and commonly used in insect molecular systematics (Yang et al. 2000). Various regions of the ribosomal repeat unit have been applied to all levels of fly divergence, and 28S rDNA is thought to be most informative for relationships as old as the Jurassic (Pélandakis et al. 1991, Pawlowski et al. 1996, Porter & Collins 1996, Friedrich & Tautz 1997 a,b, Whiting et al. 1997, Wiegmann et al. 2000b, Yang et al. 2000).

The EF-1α gene encodes elongation factor-1 alpha, an enzyme that catalyzes the GTP-dependent binding of aminoacyl-tRNA to ribosomes (Kim et al. 1990). EF-1α is among the more abundant proteins in eukaryotic cells and is expressed in almost all kinds of mammalian cells (Kim et al. 1990). EF-1α has been applied within families and subfamilies in insects (Cho et al. 1995, Mitchell et al. 1997, Yang et al. 2000). Although EF-1α has a highly conserved amino acid sequence, it retains useful phylogenetic signal in its third positions for Tertiary (65mya) and younger divergences (Cho et al. 1995, Mitchell et al. 1997, Yang et al. 2000).

Current Therevidae Classification

Therevidae or stiletto flies are a medium-sized (~1000 described species) family of flies that likely originated 150-200 mya in the Jurassic period (Hennig 1981, Evenhuis 1994). Current classification separates species of Therevidae into three subfamilies: Phycinae, Therevinae (Lyneborg 1976; Irwin & Lyneborg 1981a) and Agapophytinae (Winterton et al. 2001). Phycinae have been sub-divided into two tribes, Phycini and Xestomyzini (Lyneborg 1976), Therevinae includes the tribe Cyclotelini (Gaimari & Irwin 2000), but a tribal-level classification has not been proposed for Agapophytinae. Information about therevid biology, behavior, and classification can be found in the following publications: Lundbeck (1908), Collinge (1909), Verall (1909), Cole (1923, 1969), and English (1950), Irwin (1971, 1972, 1973, 1976, 1977, 2001), Irwin & Lyneborg (1981a, 1981b, 1989), Irwin & Yeates (1998), Winterton & Irwin (1999), and Winterton et al. (1999, 2000, 2001), Yang et al. (2000).

Sequences from opsin, CAD, 28S rDNA, and EF-1α were analyzed with equally weighted parsimony methods in the following ways: first, a combined analysis of 28S rDNA and EF-1α with and without third positions in EF-1α; second, a combined analysis of opsin and CAD with and without third positions; third, a combined analysis of 28S rDNA, EF-1α, opsin, and CAD with and without third positions. Finally, sequences from opsin, CAD, 28S rDNA, and EF-1α were also combined and analyzed using Bayesian tree reconstruction methods.

Analyses of these nucleotide sequences resulted in: 1) the production of 26 new sequences of CAD in Diptera; 2) an analysis of a subset of the Yang et al. (2000) 28S rDNA and EF-1α dataset, including some newly sequenced, key stiletto fly taxa; 3) an analysis of relationships within the Therevidae with two new nuclear genes, opsin and CAD; 4) the production of the most comprehensive molecular phylogenetic analysis of relationships within the Therevidae to date, using both parsimony and Bayesian methods for analysis of the combined DNA dataset of 28S rDNA, EF-1α, opsin, and CAD; 5) further evidence supporting the major

outline of the current therevid classification; and 6) the conclusion that the Xestomyzini tribe (Phycinae) may need to be elevated to it's own subfamily.

MATERIALS AND METHODS

Taxon Sampling

We obtained nucleotide sequences from a total of 22 genera of Therevidae, including the three recognized subfamilies, Phycinae and Therevinae (Lyneborg 1976; Irwin & Lyneborg, 1981a) and Agapophytinae (Winterton et al. 2001) (Table 1). Two representative taxa from Asiloidea were sequenced: *Mythicomyia* (Bombyliidae) and *Stenomphrale teutankhameni* (Scenopinidae). All specimens were collected directly into 95% ethanol and stored at -80ºC in the laboratory. We have opsin, CAD, 28S rDNA, and EF-1α sequences for a total of 26 taxa. The majority of the sequences for 28S rDNA and EF-1α were obtained from the previous work of Yang et al. (2000).

Laboratory methods

Total genomic nucleic acids were extracted using the guanidinium isothiocyanate homogenization protocol of Chirgwin et al. (1979). Nucleic acid pellets were resuspended in 50µl TE buffer and stored at -80˚C. Whole specimens or partial specimens (i.e.: heads, wings, and legs) were stored in 95% ethanol at the Department of Entomology of North Carolina State University Insect Collection Genome Bank.

Opsin. We initially amplified an approximately 650 base pair (bp) fragment of opsin for most taxa using degenerate primers (ops102f/ops336r; Table 2). These primers were based on an alignment of published dipteran opsin sequences and one newly characterized therevid opsin sequence. The majority of products from the initial amplification of opsin were then re-amplified using more specific primers (ops1f/ops3r; Table 2) based on the available therevid sequence, resulting in an approximately 350bp fragment or when possible they were re-amplified with ops1f/ops336r to get a larger fragment (~700bp).

Amplification with ops102f/ops336r followed a touchdown polymerase chain reaction

(PCR) protocol as follows: initial 94˚C for 4 min followed by 5 cycles of 94˚C for 30 sec, 48˚C for 30 sec, 72˚C for 1 min and then 37 cycles of 94˚C for 30 sec, 43˚C for 30 sec, 72˚C for 1 min followed by an additional elongation step of 72˚C for 3 min. Amplified products (2 lanes of 20µl per taxon) were visualized and cut from a 1% low melt gel; 100µl of depc water was then added to the gel slices and melted in a 70ºC water bath for 2 min. Gel cuts were re-amplified using the following PCR protocol: initial 94˚C for 5 min followed by 36 cycles of 94ºC for 1 min, 42ºC for 1 min, 72ºC for 1 min followed by an additional elongation step of 72ºC for 5 min. Products from re-amplification (1 lane of 25µl per taxon) were then visualized and cut from a 1% low melt gel and purified using the Qiaquick® Gel Extraction Kit (Qiagen, Inc., Santa Clara, CA). Purified products were directly sequenced using d-Rhodamine Terminator Cycle Sequencing Ready Reaction with AmpliTaqFS® DNA polymerase (PE Applied Biosystems, Foster City, CA) following the suggested protocol except annealing temperature used was 40ºC instead of the suggested 50ºC for improved results. Cycle sequencing reactions used the same primers that were used in the re-amplification PCR protocols listed in Table 2.

CAD. We amplified and sequenced CAD in two overlapping fragments with multiple combinations of primers (Table 2) designed by J. K. Moulton (UTN). The first approximately 1100bp fragment was amplified using the primer pair CAD320F/CAD691R or CAD320F/680R. If necessary, the resulting PCR products were re-amplified using one of the following primer pairs: CAD320F/CAD680R; CAD320F/CAD654R; CAD329/CAD654; CAD338/CAD680, producing an approximately 1000bp fragment. The second piece of CAD was amplified with CAD581replF/CAD843replR, which resulted in an approximately 1000bp fragment and was either directly sequenced or was re-amplified using CAD581replFCAD843R or CAD581replF/CAD835R. The resulting re-amplification PCR product was approximately 1000bp long.

Amplification with CAD320F/CAD691R or CAD320F/680R followed a touchdown PCR protocol as follows: initial 94˚C for 4 min followed by 5 cycles of 94˚C for 30 sec, 52˚C for 30

sec, 72˚C for 1 min and then 5 cycles of 94˚C for 30 sec, 48˚C for 30 sec, 72˚C for 1 min and then 37 cycles of 94˚C for 30 sec, 43˚C for 30 sec, 72˚C for 1 min followed by an additional elongation step of 72˚C for 3 min. As necessary, amplified products (2 lanes of 20µl per taxon) were visualized and cut from a 1% low melt gel, 100µl of depc water was added to the gel slices and melted in a 70ºC water bath for 2 min. Gel cuts were re-amplified using the following the same touchdown PCR protocol used with amplification of the original products.

Amplification with CAD581replF/CAD843replR followed a touchdown PCR protocol as follows: initial 94˚C for 4 min followed by 5 cycles of 94˚C for 30 sec, 53˚C for 30 sec, 72˚C for 1 min then 37 cycles of 94˚C for 30 sec, 47˚C for 30 sec, 72˚C for 1 min followed by an additional elongation step of 72˚C for 3 min. As necessary, amplified products (2 lanes of 20µl per taxon) were visualized and cut from a 1% low melt gel, 100µl of depc water was added to the gel slices and melted in a 70ºC water bath for 2 min to purify the products. Gel cuts were re-amplified using the following the same touchdown PCR protocol used with amplification of the original products. Products from re-amplification (1 lane of 25µl per taxon) were then visualized and cut from a 1% low melt gel and purified using the Qiaquick® Gel Extraction Kit (Qiagen, Inc., Santa Clara, CA).

Purified products were directly sequenced using d-Rhodamine Terminator Cycle Sequencing Ready Reaction with AmpliTaqFS® DNA polymerase (PE Applied Biosystems, Foster City, CA) following the suggested protocol except annealing temperature used was 40ºC instead of the suggested 50ºC for improved results. Cycle sequencing reactions used the same primers that were used in the re-amplification PCR protocols listed in Table 2.

28S rDNA. We amplified 28S rDNA with primer pair rc28H/28Z, which were developed by Hamby et al. (1988) and modified for use in therevids and related dipterans by Yang et al. (2000) (Table 2). Amplification followed a standard three-step PCR protocol with a 50ºC annealing temperature and 30 cycles, resulting in an approximately 1200bp fragment. PCR

products were visualized on a 1% agarose gel and purified using Qiaquick® PCR Purification Kits (Qiagen, Inc., Santa Clara, CA). Purified products were directly sequenced with the following primer pairs: rc28H/28K and rc28Q/28Z using d-Rhodamine Terminator Cycle Sequencing Ready Reaction with AmpliTaqFS® DNA polymerase (PE Applied Biosystems, Foster City, CA) following the suggested protocol.

EF-1α. Amplification of EF-1α was accomplished through standard and touchdown PCR from DNA templates and RT-PCR of mRNA transcripts according to protocols of Kawasaki (1990). EF-1α 5' PCR and EF-1α 3' PCR (developed by Kevin Moulton, UTN; Table 2) were used to amplify over 1000bp. Amplification followed a touchdown PCR protocol as follows: initial 94˚C for 4 min followed by 5 cycles of 94˚C for 30 sec, 52˚C for 30 sec, 72˚C for 1 min and then 5 cycles of 94˚C for 30 sec, 47˚C for 30 sec, 72˚C for 1 min and then 37 cycles of 94˚C for 30 sec, 42˚C for 30 sec, 72˚C for 1 min followed by an additional elongation step of 72˚C for 3 min. Amplified products (2 lanes of 20µl per taxon) were gel purified by visualization and cutting from a 1% low melt gel, adding 100µl of depc water to the gel slices and melting in a 70ºC water bath for 2 min.

Re-amplification of EF-1α was accomplished with primer pairs: EF-1α 5' PCR/Shemp and Joe-2/ EF-1α 3' PCR and followed a standard PCR protocol with an annealing temperature of 42ºC and 35 cycles. Products from re-amplification (1 lane of 25µl per taxon) were gel purified by visualization and cutting from a 1% low melt gel and purified using the Qiaquick® Gel Extraction Kit (Qiagen, Inc., Santa Clara, CA).

RT-PCR was used for other taxa that would not amplify using touchdown and standard PCR protocols. Primers EF4 and EF5 were used in the RT-PCR and the resulting 1100bp products (2 lanes of 20µl per taxon) were visualized and cut from a 1% low melt gel, 100µl of depc water was added to the gel slices and melted in a 70ºC water bath for 2 min. These gel cuts

were then re-amplified with primer pairs EF4/EF6 and EF2/EF5 and purified using the Qiaquick® Gel Extraction Kit (Qiagen, Inc., Santa Clara, CA).

Purified products of EF-1α were directly sequenced using d-Rhodamine Terminator Cycle Sequencing Ready Reaction with AmpliTaqFS® DNA polymerase (PE Applied Biosystems, Foster City, CA) following the suggested protocol except annealing temperature used was 42ºC instead of the suggested 50ºC for improved results. Cycle sequencing reactions used the same primers that were used in the re-amplification PCR protocols listed in Table 2.

Oligonucleotide primers (Table 2) designed to amplify DNA were synthesized by Genosys Biotechnologies, Inc. (Woodlands, Texas). All PCR protocols used TaKaRa Ex Taq™ Hot Start Version (TaKaRa Biomedicals, Japan). We sequenced sense and antisense DNA strands for confirmation of all sequences. Sequences were gel fractionated and base-called on an ABI PRISM 377 DNA automated sequencer (PE Applied Biosystems, Foster City, CA).

Nucleotide Alignment and Phylogenetic Analysis

Therevid sequence editing, contig assembly, and consensus calculations were performed in Sequencher 4.1 (Gene Codes Corp., Ann Arbor, MI) for the Power Macintosh. Consensus sequences were aligned manually using the multiple alignment editor of Genetic Data Environment (GDE 2.2, Smith et al. 1994). Gaps in aligned regions were treated as missing data.

Introns in opsin were identified by comparing therevid opsin sequences to known insect opsin sequences (Carulli et al. 1994), and were excluded from the data set prior to phylogenetic analysis due to their considerable sequence divergence and resulting alignment instability. Excision of introns resulted in a fragment ~300bp long. Manual alignment of the nucleotides and their corresponding amino acids of therevid opsin sequences were based on the published Carulli et al. alignment (1994).

Introns in CAD were identified by the presence of stop codons in the amino acid alignment and were excised based on the characteristic GT at the beginning and AG at the end of the intron. Excision of introns resulted in a fragment ~1500bp long. Manual alignment of the

nucleotides and their corresponding amino acids were unambiguous due to reasonably conserved regions. Alignment of additional 28S rDNA and EF-1α nucleotide sequences was based on the previously published alignment of Yang et al. (2000).

The topic of character congruence and whether or not to combine data sets from disparate sources has been intensely debated in systematics (Mickevich & Farris 1981, Miyamoto 1981, Swofford 1991, Bull et al. 1993, Chippindale & Wiens 1994, Cunningham 1997, Graham et al. 1998, Wenzel & Siddall 1999, Dolphin et al. 2000, Yoder et al. 2001, Barker & Lutzoni 2002, Darlu & Lecointre 2002, Dowton & Austin 2002, Lambkin et al. 2002). Recent empirical and simulation studies show serious potential flaws in the combinability tests (Graham et al. 1998, Wenzel & Siddall 1999, Dolphin et al. 2000, Yoder et al. 2001, Barker & Lutzoni 2002, Darlu & Lecointre 2002, Dowton & Austin 2002). In light of these studies, we used the total evidence approach with our data set, where all four genes were used in various combinations to reconstruct stiletto fly phylogeny.

Equally Weighted Parsimony. Maximum parsimony (MP) analyses were implemented in PAUP* version 4.0 using the heuristic search option and 20 replicate random addition searches with tree-bisection reconnection (TBR) branch swapping. In all MP analyses the following settings were used: all characters are of type unordered, all characters have equal weight, gaps are treated as missing, multistate taxa treated as uncertainty, steepest descent option not in effect, MaxTrees unlimited (auto-increased by 100), branches collapsed if maximum branch-length is zero, and MulTrees option in effect. When heuristic searches resulted in multiple equally parsimonious trees (EPTs), we summarized them in a strict consensus tree. Branch support estimates for most parsimonious trees (MPTs) and EPTs were calculated with nonparametric bootstrap analysis (Felsenstein 1985). Bootstrap analyses were performed with 1000 pseudoreplications of random addition sequences and TBR branch swapping. We calculated Bremer support values for the topology from the combined data using the program TreeRot v.2 (Sorenson 1999) in combination with PAUP*. TreeRot creates a NEXUS command file which is

then executed in PAUP* where a heuristic search is executed with 20 random addition replications for each constrained node in the search tree. MaxTrees was set to unlimited during Bremer support calculations.

Bayesian Analysis. Bayesian phylogenetic analyses were conducted with MrBayes 2.0 (Huelsenbeck & Ronquist 2001) using the GTR + I model of evolution. Bayesian analyses were initiated with random starting trees and were run for 1.0×10^6 generations. Sampling the Markov chain at intervals of 10,000 generations reduced the data to 100 sampling points.

An integral facet of Bayesian analysis is to ensure that the Markov chain has reached a "plateau or leveling off point." All of the sample points prior to this "plateau" are considered to be random and should be discarded as "burn-in" samples because they do not contain useful information about the parameters. We compared the log likelihood scores of sample points against the generation times and determined that a plateau was reached when the sample points came to a stable equilibrium value (Huelsenbeck & Ronquist 2001). To ensure that our analyses were not trapped on local optima, we ran four independent analyses, each beginning with different starting trees. The stable equilibrium values for each analysis were then compared for convergence. Independent analyses were considered to have converged if their log likelihood values approached similar mean values.

After discarding burn-in samples and evaluating convergence, the remaining samples were retained for further analysis. Each sample has a tree topology that includes branch lengths and substitution model parameter values. The topologies were used to generate a 50% majority rule consensus tree, with the percentage of samples recovering any particular clade representing that clade's posterior probability (Huelsenbeck & Ronquist 2001). These are considered to be the true probabilities of the clades under the assumed models (Rannala & Yang 1996), and thus we consider any probabilities of 95% and greater to be significantly supported.

RESULTS

Phylogenetic Analyses

28S rDNA and EF-1α. The combined data for the 26 taxa from 24 genera included in this analysis comprises 2258 sites, of which 1263 are 28S rDNA and 995 are EF-1α. Of these sites, 566bp were variable (25%) and 377bp phylogenetically informative (17%) (Table 3). The combined equally weighted MP analysis of 28S rDNA and EF-1α resulted in 12 EPTs of tree length 1640 with a consistency index (CI) of 0.469 and a retention index (RI) of 0.483 (Table 4). The strict consensus (Fig. 3) of the 12 trees shows a paraphyletic Therevidae, with the phycine taxa, *Hemigephyra* and *Henicomyia*, falling outside of the Therevidae. Also, *Stenomphrale teutankhameni* (Scenopinidae) is placed within the Therevidae and is a sister-group to the remaining phycine taxa with a 68% bootstrap support value (BSV). Partitioned Bremer support values (PBSVs) for this node from EF-1α and 28S rDNA seem to be in direct conflict with each other, with a value of 7 for EF-1α and –5 for 28S rDNA. The strict consensus shows very little resolution but there are several clusters that loosely represent the three known subfamilies within Therevidae. Overall, there is low bootstrap support and the PBSVs corroborate that this tree topology could change with only a few extra steps at the majority of the nodes.

The combined data for the same dataset excluding third nucleotide positions comprises 1927 sites, of which 1263 are 28S rDNA and 664 are EF-1α. Of these sites, 276bp were variable (14%) and 134bp were phylogenetically informative (7%) (Table 3). The equally weighted MP analysis excluding third nucleotide positions resulted in 145 EPTs of tree length 500 with a CI of 0.658 and a RI of 0.616 (Table 4). The strict consensus (Fig. 4) of 145 trees shows very little resolution but does reconstruct a monophyletic Therevidae (78% BSV). The Phycinae appear basal to the rest of the Therevidae, with the two taxa comprising the tribe Xestomyzini (*Hemigephyra* and *Henicomyia*) forming a basal sister-group to the rest of the phycines. The Agapophytinae form an unresolved cluster that is basal to the Therevinae. Overall, exclusion of

third nucleotide positions appears to improve the recovery of the deeper divergences (i.e. monophyly of Therevidae, subfamilies, and tribes), but decreases the resolution of more recent divergences, as seen by the numerous polytomies among closely related terminal taxa.

Opsin + CAD. The combined data for the 26 species from 24 genera included in this analysis comprises 1821sites, of which 312 are opsin and 1509 are CAD. Of these sites, 959bp were variable (53%) and 767bp were phylogenetically informative (42%) (Table 3). The combined equally weighted MP analysis of opsin and CAD resulted in nine EPTs, which are summarized in a strict consensus tree (length = 4504; CI = 0.366; RI = 0.383; Fig. 5; Table 4). The combined analysis of opsin and CAD supports a paraphyletic Therevidae (BSV < 50%), with the phycine tribe Xestomyzini (*Hemigephyra* and *Henicomyia*) forming a separate, basal clade to the rest of the Therevidae. Opsin and CAD provide insufficient variation to resolve some of the deeper divergences (i.e. the monophyly of Therevidae and the three subfamilies: Phycinae, Therevinae, and Agapophytinae). The Agapophytinae and Therevinae form a monophyletic clade with agapophytine taxa basal. Opsin and CAD resolve the Xestomyzini tribe as well as the generic relationships (*Anabarhynchus* spp. and *Ataenogera* spp.). Bootstrap support is generally low for the majority of clades.

The combined data for the same dataset excluding third nucleotide positions comprises 1214 sites, of which 208 are opsin and 1006 are CAD. Of these sites, 374bp were variable (31%) and 203bp were phylogenetically informative (17%) (Table 3). The equally weighted MP analysis excluding third nucleotide positions resulted in 15 EPTs of tree length 920 with a CI of 0.528 and a RI of 0.471 (Table 4). The strict consensus (Fig. 6) of 15 trees shows very little resolution but does reconstruct a monophyletic Therevidae (96% BSV). The relationships among the taxa within the Phycinae, Agapophytinae, and Therevinae are mostly unresolved, with low bootstrap values. Within the Phycinae, the tribe Xestomyzini (*Hemigephyra* and *Henicomyia*) and the genus *Ataenogera* are well resolved (BSVs = 100% and 99%, respectively). Within the Agapophytinae, *Entesia* and *Melanotherva* form a monophyletic group (BSV = 99%). The

therevine genus *Anabarhynchus* forms a monophyletic group with 97% bootstrap support, *Brachylinga sp*. and *Penniverpa* also form a monophyletic group (57% BSV). Exclusion of third positions decreased the overall resolution of the tree, as well as bootstrap support.

Maximum Parsimony Analysis of 28S rDNA, EF-1α, Opsin, and CAD. The combined data for the 26 taxa from 24 genera included in this analysis comprises 4079 sites, of which 1263 are 28S rDNA, 995 are EF-1α, 312 are opsin, and 1509 are CAD. Of these sites, 1525bp were variable (37%) and 1144bp were phylogenetically informative (28%) (Table 3). The combined equally weighted MP analysis of 28S rDNA, EF-1α, opsin, and CAD resulted in a single MPT, which is summarized in Fig. 7 (length = 6199; CI = 0.390; RI = 0.401; Table 4). Therevidae is monophyletic (BSV = 69%), with a paraphyletic Phycinae comprising the tree's basal lineages. The Xestomyzini (*Hemigephyra* and *Henicomyia*) form a monophyletic clade (98% BSV) separate from the rest of the Phycinae (89% BSV). The resolution of the lower level relationships within Phycinae is much improved over the analyses that included only 28S rDNA/ EF-1α and opsin/CAD. *Neotabuda* + *Ruppellia* are basal (95% BSV) to the rest of the phycine clade. The *Ataenogera* spp. clade (100% BSV) is sister to *Phycus* + *Pherocera* clade (82% BSV) with 69% bootstrap support. The paraphyletic Therevinae appear basal to the Agapophytinae. The *Anabarhynchus* genus (Therevinae) falls outside of the Therevinae and appears as a sister-group to the Agapophytinae. The lower-level relationships within the Therevinae and Agapophytinae are also well resolved with mostly high bootstrap support. As might be expected for increased dataset size, the combined analysis of all four genes gave better overall resolution throughout the tree and higher bootstrap support per node.

The combined data for the same dataset excluding third nucleotide positions comprises 3141sites, of which of which 1263 are 28S rDNA, 664 are EF-1α, 208 are opsin, and 1006 are CAD. Of these sites, 650bp were variable (21%) and 337bp were phylogenetically informative (11%) (Table 3). The equally weighted MP analysis excluding third nucleotide positions resulted

in 12 EPTs of tree length 1435 with a CI of 0.568 and a RI of 0.510 (Table 4). The strict consensus (Fig. 8) of the 12 trees is concordant with the results of Yang et al. (2000) and current therevid classifications. The Therevidae are monophyletic with 100% bootstrap support. The Phycinae are paraphyletic, with the Xestomyzini tribe (*Hemigephyra* and *Henicomyia*) forming a basal clade (100% BSV). This may suggest that the Xestomyzini tribe should be elevated to a subfamily, though increased taxonomic sampling and phylogenetic evaluation would be required before such a determination. The remaining phycine taxa form a monophyletic clade (96% BSV) with a slight rearrangement of the taxa. *Phycus* + *Ruppellia* form a monophyletic clade with fairly low bootstrap support (55%) versus when all nucleotide positions were included, *Phycus* + *Pherocera* are a monophyletic clade with 82% bootstrap support.

The Agapophytinae form a monophyletic group that is basal to the Therevinae (< 50% BSV). The lower-level classification within the Agapophytinae is fairly well resolved: *Acraspisa* sp. + *Agapophytus* (59% BSV); *Actenomeros* + New Genus S (68% BSV); and *Entesia* + *Melanothereva* (100% BSV). The Therevinae are monophyletic (BSV = 73%), with the *Anabarhynchus* spp. basal. The remaining taxa of the Therevinae are mostly unresolved with the exception of the *Brachylinga* + *Penniverpa* clade (BSV = 78%).

Exclusion of third nucleotide positions from the combined four gene-dataset resulted in the topology with greatest concordance to previous stiletto fly classifications. Bootstrap support is fairly strong for both higher and lower level divergences, with the exception of some nodes within the Therevinae.

Bayesian Analysis of 28S rDNA, EF-1α, Opsin, and CAD. Bayesian analysis of the total combined data under the GTR + I model for 1.0×10^6 generations resulted in a posterior probability distribution containing 1.0×10^2 samples per analysis. Independent analyses converged on similar log likelihood values and reached a leveling off point at 20,000 generations. The initial 20,000 samples from each analysis were discarded and the remaining combined

samples were used to compute the 50% majority rule tree seen in Fig. 9, with 20 of the 23 ingroup nodes with a significance level > 95%.

The phycine tribe Xestomyzini (*Hemigephyra* and *Henicomyia*) is again the basal-most clade, followed by the rest of the Phycinae: *Neotabuda* + *Ruppellia*, which is a sister-group to the *Ataenogera* spp. and the *Phycus* + *Pherocera.* The Agapophytinae are a monophyletic group basal to the Therevinae. Within the Agapophytinae, the *Acraspisa* + *Agapophytus* clade is sister to New Genus M. That clade is sister to the *Actenomeros* and *Entesia* + *Melanothereva* clade. Within the Therevinae, the *Anabarhynchus* genus group is basal to the rest of the therevine taxa. *Brachylinga* + *Cyclotelus pictipennis* and *Hoplosathe frauenfeldi* + *Thereva* form sister-groups.

This tree is very similar topologically to the MP analyses of the same data. The Bayesian tree is better resolved and has higher statistical support than the MP trees. The Bayesian tree is also highly concordant with current stiletto fly classifications (Yang 2000; Yang et al. 2000; Metz 2002).

DISCUSSION

Analyses of these nucleotide sequences resulted in: 1) the production of 26 new sequences of CAD in Diptera; 2) an analysis of a subset of the Yang et al. (2000) 28S rDNA and EF-1α dataset, including some newly sequenced, key stiletto fly taxa; 3) an analysis of relationships within the Therevidae with two new nuclear genes, opsin and CAD; 4) the production of the most comprehensive molecular phylogenetic analysis of relationships within the Therevidae to date, using both parsimony and Bayesian methods for analysis of the combined DNA dataset of 28S rDNA, EF-1α, opsin, and CAD; 5) further evidence supporting the major outline of the current therevid classification; and 6) the conclusion that the Xestomyzini tribe (Phycinae) may need to be elevated to it's own subfamily.

28S rDNA and EF-1α. Yang et al.'s (2000) molecular phylogenetic analysis of Therevidae supported the monophyly of the family and two major subfamilial clades: Phycinae and Therevinae (Irwin & Lyneborg 1981a). Their most reliable tree (Fig. 1) was reconstructed

from a combined analysis of 28S rDNA and EF-1α and was well-supported with high BSVs and PBSVs. However, their analyses were unable to satisfactorily resolve the relationships within the Therevinae.

Our study used a subset of the Yang et al. (2000) data and also included eight new taxa for the same genes. Maximum parsimony analyses of this new dataset were not fully concordant with the results of Yang et al. (2000). However, caution must be used when comparing the two analyses for the following reasons: 1) our dataset represents only a subset of the taxa that were used in the Yang et al. (2000) study, which includes a smaller outgroup sampling; and 2) we included a few (~8) new taxa that were not included in the Yang et al. (2000) analysis. We did not include all of the taxa that Yang et al. (2000) included due to the lack of an overlapping dataset for opsin and CAD.

Our analyses of 28S rDNA and EF-1α with all nucleotide positions included resulted in a paraphyletic Therevidae, with the phycine taxa *Hemigephyra* and *Henicomyia* falling outside of the Therevidae and the outgroup taxon *Stenomphrale teutankhameni* (Scenopinidae) forming a sister-group with the remaining phycine taxa (Fig. 3). There are two clusters of taxa that correspond with Agapophytinae and Therevinae, but overall bootstrap support was low.

Exclusion of third positions resulted in a monophyletic Therevidae, but with very little additional resolution within the family (Fig. 4). Loss of resolution may be best explained by loss of information- only 7% of the remaining nucleotide sites were phylogenetically informative (17% of sites were phylogenetically informative when third positions were included). Thus, third positions in EF-1α appear to contain more than half of the phylogenetic signal, which is consistent with the findings of Cho et al. (1995) and Mitchell et al. (1997). The PBSVs also indicate that the majority of phylogenetic signal for the major clades is coming from 28S rDNA (PBSVs for 28S rDNA are all almost double those for EF-1α). Previous studies using 28S rDNA suggest that it is most useful for older divergences, such as the late Jurassic (Whiting et al. 1997,

Friedrich & Tautz 1997, Wiegmann et al. 2000b, Yang et al. 2000) and EF-1α appears to be more informative for Tertiary and younger divergences below the family level (Mitchell et al. 1997, Friedlander et al. 1998, Yang et al. 2000). Our tree has very little resolution in the tips for more recent divergences where EF-1α should provide the most resolution and has better resolution for the deeper divergences where 28S rDNA should provide the most phylogenetic signal.

Opsin and CAD. Phylogenetic analysis of all nucleotide positions of opsin and CAD resulted in a paraphyletic Therevidae, with the Xestomyzini tribe forming a separate basal clade from the rest of the therevid taxa (Fig. 5). Overall, there are clusters of taxa that correspond with the current subfamilial classification of Therevidae, though resolution within these clades is low. The resolution of the Xestomyzini tribe as well as the genera *Anabarhynchus* and *Ataenogera* suggest increased utility at the more recent divergence times when all nucleotide positions are analyzed. There is some resolution of more recent divergences when third positions are included in the phylogenetic analysis, which most likely corresponds to third positions in protein-encoding genes being more variable than first and second positions.

When third positions are excluded, the more closely related species (i.e. within the same tribe and genera) are still resolved with high bootstrap support, but other less closely related species lose some resolution (Fig. 6). Interestingly, PBSVs for both genes are generally higher when third positions are excluded.

Opsin and CAD are phylogenetically informative within the Therevidae, but increased taxon sampling may give a better view of their performance. Our study had somewhat limited taxon sampling within Therevidae due to time constraints, in the future it would be advantageous to gather more opsin and CAD sequences that are representative of the taxa that were used in the Yang et al. (2000) study.

Opsin is part of a multiple copy gene family and we used what we believe to be one copy (see chapter one for further discussion). There are five copies of opsin characterized in *Drosophila melanogaster* (O'Tousa et al. 1985, Cowman et al. 1986, Zuker et al. 1987, Montell

et al. 1987, Chou et al. 1996). Therefore, there are possibly other uncharacterized copies of opsin that would be potentially useful phylogenetic markers within Therevidae. The copy of opsin that we were able to amplify and sequence was only 312bp long (without introns). A longer fragment of opsin would be desirable for future phylogenetic analyses but would require more time-consuming and complex molecular techniques such as cloning. Working with a new, multiple copy gene family such as opsin requires a lot of careful primer design and experimentation with PCR protocols that may deter its use as a phylogenetic marker until it is more completely characterized. This study was a first step in learning more about opsin's utility in a new taxonomic group (Diptera), since previously it has only been used as a phylogenetic marker within the Hymenoptera (Mardulyn & Cameron 1999, Cameron & Mardulyn 2001, Ascher et al. 2001, Rokas et al. 2002).

This study has produced 26 new sequences of CAD in Therevidae. Previously, phylogenetic analysis of CAD has been limited to investigating the evolution of carbomoylphosphate synthetase (CPS) genes (van den Hoff et al. 1995, Lawson et al. 1996). There has been a recent wave of interest in the phylogenetic utility of CAD in dipteran systematics that has been in part driven by the availability of primers designed by J. K. Moulton (UTN). However, CAD has other important assets that make it an attractive phylogenetic marker; it is a single-copy, nuclear protein-encoding gene, which allows for unambiguous alignment and phylogenetic analysis. Also, CAD is quite large (~6700bp) with three domains and various regions that range from very conserved to quite variable, these attributes make CAD potentially useful for both deeper and more recent divergences (Moulton pers. com.).

In this study, we only used a small portion of the CAD gene within the first domain (CPSase). The fragment that we used contained both variable and conserved regions and was relatively easy to amplify, sequence, and align. We have generated a new dataset that will be a useful tool for future studies involving CAD in therevid systematics as well as insect systematics in general.

28S rDNA, EF-1α, Opsin, and CAD. Maximum parsimony analysis of all nucleotide positions of 28S rDNA, EF-1α, opsin, and CAD resulted in a single MPT that is well resolved and has high BSVs (Fig. 7). Therevidae is monophyletic, with the phycine tribe Xestomyzini forming a basal, separate clade from the rest of the therevid taxa. Analysis of all four genes improved the overall resolution of the tree, with a few taxa falling outside of their subfamily such as the *Anarbarhynchus* spp. The PBSVs indicate that each node is differentially supported by each of the genes, which reinforces the importance of choosing multiple genes with different evolutionary histories when reconstructing phylogenetic relationships (Wheeler et al. 1993, Dowton & Austin 2002).

The Therevinae are paraphyletic and appear basal to the Agapophytinae, which is not concordant with the results of Yang et al. (2000). However, when the same dataset was analyzed excluding third positions (Fig. 8), the Therevinae appear monophyletic and the Agapophytinae are basal to the Therevinae, which is concordant with the results of Yang et al. (2000). Exclusion of third positions appears to diminish the resolution of the lower-level relationships slightly, but increase the concordance of the overall subfamilial classification with previous therevid classifications. The relationships among taxa within the Therevinae are poorly resolved from this analysis.

Combined Bayesian analysis of all nucleotide positions (Fig. 9) resulted in a well resolved tree with high statistical support (20 of the 23 ingroup nodes with a significance level > 95%). This tree is the best hypothesis of therevid relationships of all our analyses.

Overall, our analyses support the current therevid classification scheme, with the exception of the Xestomyzini. The phycine tribe Xestomyzini forms a separate, basal clade from the rest of the Phycinae. We believe that this suggests that the Xestomyzini tribe may need to be elevated to it's own subfamily. Ongoing research by M. Hauser and M. E. Irwin (UIUC, pers. comm.) including many additional xestomyzine taxa and comprehensive morphological and molecular analyses will be critical to justify this proposed elevation in taxonomic rank.

The Bayesian analysis depicts the basal relationships with much improved resolution over the parsimony analysis of the same data. Here, we have clarified lower-level relationships within the Therevidae and improved resolution within all of the subfamilies. Our combined data results analyzed using Bayesian methods provide the greatest resolution and best support yet achieved for analyses of therevid evolutionary relationships. Future studies would benefit from the use of these methods and more comprehensive taxon sampling from the highly diverse clades of Therevidae.

CONCLUSIONS

Our analyses of 28S rDNA, EF-1α, opsin, and CAD are the most comprehensive molecular phylogenetic study of the Therevidae to date. Our combined analysis of these genes provides phylogenetic resolution that is highly concordant with the current therevid classification. We also show that within the Phycinae, the Xestomyzini is monophyletic and could represent a major therevid clade.

Furthermore, we have increased the knowledge of two new nuclear genes (opsin and CAD) phylogenetic utility in therevids. Overall, we have created a new dataset (opsin and CAD) that can be used to build a larger dataset and can also be used in conjunction with a morphological dataset that is currently being generated within the therevid PEET project. Future work with opsin in therevids should include characterization and analysis of the phylogenetic utility of the other copies of opsin. Other regions within CAD should be explored and added to this current dataset to further resolve therevid relationships.

ACKNOWLEDGMENTS

We thank L.L. Deitz and E.L. Vargo for comments on this manuscript and B. Cassel and K. Moulton for their technical support and assistance. The following people provided important specimens, data, or information that enabled us to conduct this research: M. Hauser, K. Holston, M. Irwin, G. Kampmeier, K. Moulton, E. Schlinger, S. Winterton, L. Yang, and D. Yeates. This project was supported by funds from the National Science Foundation (DEB-9977958), the North

Carolina State University College of Agriculture and Life Sciences, and North Carolina Agricultural Research Services.

REFERENCES

Abraham, D., N. Ryrholm, H. Wittzell, J.D. Holloway, M.J. Scoble, C. Löfstedt. 2001. Molecular phylogeny of the subfamilies in Geometridae (Geometroidea: Lepidoptera). Mol. Phylogenet. Evol. 20: 65-77.

Ascher, J.S., B.N. Danforth, J. Shuqing. 2001. Phylogenetic utility of the major opsin in bees (Hymenoptera: Apoidea): a reassessment. Mol. Phylogenet. Evol. 19: 76-93.

Baker, R.H., G.S. Wilkinson, R. DeSalle. 2001. Phylogenetic utility of different types of molecular data used to infer evolutionary relationships among stalk-eyed flies (Diopsidae). Syst. Biol. 50: 87-105.

Barclay, P.L., J.B.C. Findlay. 1984. Labeling of the cytoplasmic domains of ovine rhodopsin with hydrophilic chemical probes. Biochem. J. 220: 75-84.

Barker, F.K., F.M. Lutzoni. 2002. The utility of the incongruence length difference test. Syst. Biol. 51: 625-637.

Briscoe, A.D., L. Chittka. 2001. The evolution of color vision in insects. Annu. Rev. Entomol. 46: 471-510.

Buckley, T.R., P. Arensburger, C. Simon, G.K. Chambers. 2002. Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. Syst. Biol. 51: 4-18.

Bull, J.J., J.P. Huelsenbeck, C.W. Cunningham, D.L. Swofford, P.J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42: 384-397.

Cameron, S.A., P. Mardulyn. 2001. Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera: Apinae). Syst. Biol. 50: 194-214.

Carulli, J.P., D.-M. Chen, W.S. Stark, D.L. Hartl. 1994. Phylogeny and physiology of *Drosophila* opsins. J. Mol. Evol. 38: 250-262.

Caterino, M.S., R.D. Reed, M.M. Kuo, F.A.H. Sperling. 2001. A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera: Papilionidae). Syst. Biol. 50: 106-127.

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

Chippindale, P.T., J.J. Wiens. 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. Syst. Biol. 43: 278-287.

Chirgwin, J.M., A.E. Przybyla, R.J. MacDonald, W.J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18: 5294.

Cho, S., A. Mitchell, J.C. Regier, C. Mitter, R.W. Poole, T.P. Friedlander, S. Zhao. 1995. A highly conserved nuclear gene for low-level phylogenetics: elongation-factor 1α recovers morphology-based tree for heliothine moths. Mol. Biol. Evol. 12: 650-656.

Chou, W.-H., K.J. Hall, D.B. Wilson, C.V. Wideman, S.M. Townson, L.V. Chadwell, S.G. Britt. 1996. Identification of a novel *Drosophila* opsin reveals specific patterning of the R7 and R8 photoreceptor cells. Neuron 17: 1101-1115.

Cole, F.R. 1923. A revision of the North American two-winged flies of the family Therevidae. Proc. U. S. Natl. Mus. 62: 1-140.

Cole, F.R. 1969. The Flies of Western North America. University of California Press, Berkeley, California.

Collinge, W.E. 1909. Observations on the life-history and habits of *Thereva noblitata* Fabr., and other species. J. Econ. Biol. 4: 14-18.

Cowman, A.F., C.S. Zuker, G.M. Rubin. 1986. An opsin gene expressed in only one photoreceptor cell type of the *Drosophila* eye. Cell 44: 705-710.

Crandall, K.A., T.W. Cronin. 1997. The molecular evolution of visual pigments of freshwater crayfishes (Decapoda: Cambaridae). J. Mol. Evol. 45: 524-534.

Cunningham, C.W. 1997. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. Syst. Biol. 46: 464-478.

Darlu, P., G. Lecointre. 2002. When does the incongruence length difference test fail? Mol. Biol. Evol. 19: 432-437.

Dolphin, K., R. Belshaw, C.D.L. Orme, D.L.J. Quicke. 2000. Noise and incongruence: Interpreting results of the incongruence length difference test. Mol. Phylogenet. Evol. 17: 401-406.

Dowton, M., A.D. Austin. 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy- the behavior of the incongruence length difference test in mixed-models analyses. Syst. Biol. 51: 19-31.

English, K.M.I. 1950. Notes on the morphology and biology of *Anabarhynchus fasciatus* Macq. and other Australian Therevidae. Proc. Linn. Soc. N.S.W. 75: 345-359.

Evenhuis, N.L. 1994. Catalogue of the Fossil Flies of the World (Insecta: Diptera). Backhuys, Leiden.

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

Flook, P.K., S. Klee, C.H.F. Rowell. 1999. Combined molecular phylogenetic analysis of the Orthoptera (Arthropoda, Insecta) and implications for their higher systematics. Syst. Biol. 48: 233-253.

Freund, J.N., W. Vergis, P. Schedl, B.P. Jarry. 1986. Molecular organization of the rudimentary gene of *Drosophila melanogaster*. J. Mol. Biol. 189: 25-36.

Friedlander, T.P., K.R. Horst, J.C. Regier, C. Mitter, R.S. Peigler, Q.Q. Fang. 1998. Two nuclear genes yield concordant relationships within the Attacini (Lepidoptera: Saturniidae). Mol. Phylogenet. Evol. 9: 131-140.

Friedrich, M., D. Tautz. 1997a. An episodic change of rDNA nucleotide substitution rate has occurred during the emergence of the insect order Diptera. Mol. Biol. Evol. 14: 644-653.

Friedrich, M., D. Tautz. 1997b. Evolution and phylogeny of the Diptera: A molecular phylogenetic analysis using 28S rDNA sequences. Syst. Biol. 46: 674-698.

Gaimari, S.D., & M.E. Irwin. 2000. Phylogeny, classification, and biogeography of the cycloteline Therevinae (Diptera: Therevidae). Zool. J. Linn. Soc. 129: 129-240.

Graham, S.W., J.R. Kohn, B.R. Morton, J.E. Eckenwalder, S.C.H. Barrett. 1998. Phylogenetic congruence and discordance among one morphological and three molecular data sets from Pontederiaceae. Syst. Biol. 47: 545-567.

Hamby, R.K., L. Sims, L. Issel, E. Zimmer. 1988. Direct ribosomal RNA sequencing: optimization of extraction and sequencing methods for work with higher plants. Plant Mol. Biol. Rep. 6: 175-192.

Hargrave, P.A., J.H. McDowell, D.R. Curtis, J.K. Wang, E. Juszczak, S.L. Fong, J.K. Mohanna Rao, P. Argos. 1983. The structure of bovine rhodopsin. Biophys. Struct. Mech. 9: 235-244.

Hennig, W. 1973. Diptera. In "Handbuch der Zoologie, IV: Arthropoda" (W. Kukenthal, Ed.), pp. 1-337. de Gruyter, New York.

Hennig, W. 1981. Insect Phylogeny. Wiley, New York.

Huelsenbeck, J.P, F. Ronquist. 2001. MR-BAYES: Bayesian inference of phylogeny. Bioinformatics 17: 754-755.

Irwin, M.E. 1971. Ecology and Biosystematics of the Pherocerine Therevidae (Diptera: Phycinae). Ph.D. Dissertation, University of California, Riverside, California.

Irwin, M.E. 1972. Diagnosis and habitat preferences of the immature stages of three South African species of the *Xestomyza*-group (Diptera: Therevidae). Ann. Natal Mus. 21: 377-389.

Irwin, M.E. 1973. A new genus of the *Xestomyza*-group from the western coast of South Africa, based on two new species with flightless females (Diptera: Therevidae). Ann. Natal Mus. 21: 533-556.

Irwin, M.E. 1976. Morphology of the female terminalia and known ovipositing behaviour of female Therevidae (Diptera: Asiloidea), with an account of correlated adaptations and comments on phylogenetic relationships. Ann. Natal Mus. 22: 913-935.

Irwin, M.E. 1977. Two new genera and four new species of the *Pherocera*-group from western North America, with observations on habitats and behavior (Diptera: Therevidae: Phycinae). Proc. Entomol. Soc. Wash. 79: 422-451.

Irwin, M.E. 2001. Species composition and seasonal flight periodicity of stiletto flies (Diptera: Therevidae) occurring along the Kuiseb River, Gobabeb, Namibia. Cimbebasia 17: 169-175.

Irwin, M.E., G. Kampmeier. 1997. PEET: a training effort that is paying dividends. Ill. Nat. Hist. Surv. Rep. 347: 2.

Irwin, M.E., L. Lyneborg. 1981a. The genera of Nearctic Therevidae. Bull. Ill. Nat. Hist. Surv. 32: 193-277.

Irwin, M.E., L. Lyneborg. 1981b. Therevidae. In "Manual of Nearctic Diptera" (J.F. McAlpine, B.V. Peterson, G.E. Shewell, H.J. Teskey, J.R. Vockeroth, D.M. Wood, Eds.), pp. 513-523. Research Branch, Agriculture Canada Monograph 1.b.

Irwin, M.E., L. Lyneborg. 1989. Family Therevidae. In "Catalog of Diptera of the Australasian and Oceanian Regions" (N.L. Evenhuis, Ed.), pp. 353-358. Bishop Museum Special Publication 86. Bishop Museum Press, Honolulu.

Irwin, M.E., D.K. Yeates. 1998. "The Internal Hierarchy of the Therevidae and Biogeographic Patterns of Extant Higher Rank Taxa (Brachycera: Asiloidea)," Abstracts Volume, Fourth International Congress of Dipterology, p. 90. Oxford, United Kingdom.

Jaroslaw, K., R.C. Wilkerson, N.J. Besansky. 2001. Toward understanding Anophelinae (Diptera, Culicidae) phylogeny: Insights from nuclear single-copy genes and the weight of evidence. Syst. Biol. 50: 540-556.

Jarry, B.P., D. Falk. 1974. Functional diversity within the *rudimentary* locus of *Drosophila melanogaster*. Mol. Gen. Genet. 135: 113-122.

Johnson, K.P., M.F. Whiting. 2002. Multiple genes and the monophyly of Ischnocera (Insecta: Phthiraptera). Mol. Phylogenet. Evol. 22: 101-110.

Joy, D.A., J.E. Conn. 2001. Molecular and morphological phylogenetic analysis of an insular radiation in Pacific black flies (*Simulium*). Syst. Biol. 50: 18-38.

Kawasaki, E.S. 1990. Amplification of DNA. In "PCR Protocols: a Guide to Methods and Applications" (M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White, Eds.). Academic Press, San Diego.

Kim, D.W., T. Uetsuki, Y. Kaziro, N. Yamaguchi, S. Sugano. 1990. Use of the human elongation factor-1 alpha promoter as a versatile and efficient expression system. Gene 91: 217-223.

Kjer, K.M., R.J. Blahnik, R.W. Holzenthal. 2001. Phylogeny of Trichoptera (caddisflies): characterization of signal and noise within multiple datasets. Syst. Biol. 50: 781-816.

Lambkin, C.L., M.S.Y. Lee, S.L. Winterton, D.K. Yeates. 2002. Partitioned Bremer support and multiple trees. Cladistics 18: 436-444.

Lawson, F.S., R.L. Charlebois, J.R. Dillon. 1996. Phylogenetic analysis of carbomoylphosphate synthetase genes: Complex evolutionary history includes the internal duplication within a gene which can root the tree of life. Mol. Biol. Evol. 13: 970-977.

Lundbeck, W. 1908. Therevidae. In "Diptera Danica. Genera and Species of Flies Hitherto Found in Denmark. Part 2", pp. 132-155. Gad, Copenhagen.

Lyneborg, L. 1976. A revision of the therevine stiletto-flies (Diptera: Therevidae) of the Ethiopian region. Bull. Br. Mus. (Nat. Hist.) Entomol. 33: 191-346.

Majer, J.M. 1997. Family Therevidae. In "Contributions to a Manual of Palearctic Diptera, Vol. 2, Nematocera and Lower Brachycera" (L. Papp and B. Darvas, Eds.), pp. 519-529. Science Herald, Budapest.

Mardulyn, P., S.A. Cameron. 1999. The major opsin in bees (Insecta: Hymenoptera): a promising nuclear gene for higher level phylogenetics. Mol. Phylogenet. Evol. 12: 168-176.

Martynov, V.I., M.B. Kostina, M. Yu. Feigina, A.I. Miroshnikov. 1983. A study of the molecular organization of visual rhodopsin in photoreceptor membranes by limited proteinolysis. Bioorg. Khim. 9: 734-745.

Metz, M.A. 2002. Systematics within the Therevinae (Insecta: Diptera: Therevidae). Ph.D. Dissertation. Department of Entomology, University of Illinois at Urbana-Champaign.

Mickevich, M., J. Farris. 1981. The implications of congruence in *Menidia*. Syst. Zool. 30: 351-370.

Mitchell, A., S. Cho, J.C. Regier, C. Mitter, R.W. Poole, M. Matthews. 1997. Phylogenetic utility of elongation factor-1α in Noctuoidea (Insecta: Lepidoptera): the limits of synonymous substitution. Mol. Biol. Evol. 14: 381-390.

Miyamoto, M.M. 1981. Congruence among character sets in phylogenetic studies of the frog genus *Leptodactylus*. Syst. Zool. 30: 281-290.

Montell, C., K. Jones, C. Zuker, G. Rubin. 1987. A second opsin gene expressed in the ultraviolet-sensitive R7 photoreceptor cells of *Drosophila melanogaster*. J. Neurosci. 7: 1558-1566.

Mullen, E., M. Akhtar. 1981. Topographic and active sites studies on bovine rhodopsin. FEBS Lett. 132: 261-264.

Nathans, J., D.S. Hogness. 1983. Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. Cell 34: 807-814.

Nathans, J., D.S. Hogness. 1984. Isolation and nucleotide sequence of the gene encoding human rhodopsin. Proc. Natl. Acad. Sci. U.S.A. 81: 4851-4855.

Nørby, S. 1973. The biochemical genetics of rudimentary mutants of *Drosophila melanogaster*. I. Aspartate carbamoyltransferase levels in complementing and non-complementing strains. Hereditas 73: 11-16.

O'Tousa, J., W. Baehr, R.L. Martin, J. Hirsch, W.L. Pak, M.L. Applebury. 1985. The *Drosophila ninaE* gene encodes an opsin. Cell 40: 839-850.

Ovchinnikov, Yu. A. 1982. Rhodopsin and bacteriorhodopsin: structure-function relationships. FEBS Lett. 148: 179-191.

Ovchinnikov, Yu. A., N.G. Abdulaev, M. Yu. Feigina, I.D. Artamonov, A.S. Bogachuk, A.S. Zolotarev, E.R. Eganyan, P.V. Kostetskii. 1983. Visual rhodopsin III: complete amino acid sequence and topography in the membrane. Bioorg. Khim. 9: 1331-1340.

Pawlowski, J., R. Szadziewski, D. Kmieciak, J. Fahrni, G. Bittar. 1996. Phylogeny of the infraorder Culicomorpha (Diptera: Nematocera) based on 28S rRNA gene sequences. Syst. Entomol. 21: 167-178.

Pélandakis, M., D.G. Higgins, M. Solignac. 1991. Molecular phylogeny of the subgenus *Sophophora* of *Drosophila* derived from large subunit of ribosomal RNA sequences. Genetica 84: 87-94.

Porter, C.H., F.H. Collins. 1996. Phylogeny of Nearctic members of the *Anopheles maculipennis* species group derived from the D2 variable region of 28S ribosomal RNA. Mol. Phylogenet. Evol. 6: 178-188.

Rannala, B., Z. Yang. 1996. Probability distribution of molecular evolutionary trees. J. Mol. Evol. 43: 304-311.

Rawls, J.M., J.W. Fristrom. 1975. A complex genetic locus that controls the first three steps of pyrimidine biosynthesis in *Drosophila*. Nature 255: 738-740.

Rawls, J.M., L.A. Porter. 1979. Organization of the *rudimentary wing* locus in *Drosophila melanogaster*. I. The isolation and partial characterization of mutants induced with ethyl methanesulfonate, ICR-170, and x rays. Genetics 93: 143-161.

Rokas, A., J.A.A. Nylander, F. Ronquist, G.N. Stone. 2002. A maximum-likelihood analysis of eight phylogenetic markers in gallwasps (Hymenoptera: Cynipidae): implications for insect phylogenetic studies. Mol. Phylogenet. Evol. 22: 206-219.

Simmons, A.J., J.M. Rawls, J. Pikur, J.N. Davidson. 1999. A mutation that uncouples allosteric regulation of carbamyl phosphate synthetase in *Drosophila*. J. Mol. Biol. 287: 277-285.

Smith, S.W., R. Overbeek, C.R. Woese, W. Gilbert, P.M. Gillevet. 1994. The genetic data environment and expandable GUI for multiple sequence analysis. Comput. Appl. Biosci. 10: 671-675.

Sorenson, M.D. 1999. TreeRot, version 2. Boston University, Boston, MA.

Swofford, D. 1991. When are phylogeny estimates from molecular and morphological data incongruent? In "Phylogenetic Analysis of DNA Sequences" (M.M. Miyamoto, J. Cracraft, Eds.), pp. 295-333. Oxford University Press, New York.

Swofford, D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer, Sunderland, MA.

van den Hoff, M.J.B., A. Jonker, J.J. Beintema, W.H. Lamers. 1995. Evolutionary relationships of the carbomoylphosphate synthetase genes. J. Mol. Evol. 41: 813-832.

Verrall, G.H. 1909. Stratiomyidae and succeeding families of the Diptera Brachycera of Great Britain. In "British Flies, Vol. 5", (G.H. Verrall, Ed.), pp. 1-780. Gurney & Jackson, London.

Weekers, P.H.H., J.F. De Jonckheere, H.J. Dumont. 2001. Phylogenetic relationships inferred from ribosomal ITS sequences and biogeographic patterns in representatives of the genus *Calopteryx* (Insecta: Odonata) of the west Mediterranean and adjacent west European zone. Mol. Phylogenet. Evol. 20: 89-99.

Wenzel, J.W., M.E. Siddall. 1999. Noise. Cladistics 15: 51-64.

Wheeler, W. C., P. Cartwright, C.Y. Hayashi. 1993. Arthropod phylogeny: A combined approach. Cladistics 9: 1-39.

Whiting, M.F., J.C. Carpenter, Q.D. Wheeler, W.C. Wheeler. 1997. The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46: 1-68.

Wiegmann, B.M., C. Mitter, J.C. Regier, T.P. Friedlander, D.M. Wagner, E.S. Nielsen. 2000a. Nuclear genes resolve Mesozoic-aged divergences in the insect order Lepidoptera. Mol. Phylogenet. Evol. 15: 242-259.

Wiegmann, B.M., S.C. Tsaur, D.W. Webb, D.K. Yeates, B.K. Cassel. 2000b. Monophyly and relationships of the Tabanomorpha (Diptera: Brachycera) based on 28S ribosomal gene sequences. Ann. Entomol. Soc. Am. 93: 1031-1038.

Wiegmann, B.M., J.C. Regier, C. Mitter. 2001. Combined molecular and morphological evidence on the phylogeny of the earliest lepidopteran lineages. Zool. Scr. 31: 67-81.

Winterton, S.L., M.E. Irwin. 1999. *Laxotela*- a new genus of Therevidae (Diptera) from Australia. Ent. scand. 30: 299-310.

Winterton, S.L., M.E. Irwin, D.K. Yeates. 1999. Systematics of *Nanexila* Winterton & Irwin, gen. Nov. (Diptera: Therevidae) from Australia. Invertebr. Taxon. 13: 237-308.

Winterton, S.L., J.H. Skevington, M.E. Irwin, D.K. Yeates. 2000. Phylogenetic revision of *Bonjeania* Irwin & Lyneborg (Diptera: Therevidae). Syst. Entomol. 25: 295-324.

Winterton, S.L., L. Yang, B.M. Wiegmann, D.K. Yeates. 2001. Phylogenetic revision of Agapophytinae subf.n. (Diptera: Therevidae) based on molecular and morphological evidence. Syst. Entomol. 26: 173-211.

Woodley, N.E. 1989. Phylogeny and classification of the "Orthorrhaphous" Brachycera. In "Manual of the Nearctic Diptera, Vol. 3" (J.F. McAlpine, Ed.), pp. 1371-1395. Monograph No. 32, Research Branch, Agriculture Canada, Ottawa.

Yang, L.L. 2000. Molecular phylogenetics of the Therevidae and their position among the families of the Asiloidea (Insecta: Diptera). Ph.D. Thesis. Department of Entomology, North Carolina State University.

Yang, L.L., B.M. Wiegmann, D.K. Yeates, M.E. Irwin. 2000. Higher-level phylogeny of the Therevidae (Diptera: Insecta) based on 28S Ribosomal and Elongation Factor-1_ gene sequences. Mol. Phylogenet. Evol. 15: 440-451.

Yeates, D.K. 1994. The cladistics and classification of the Bombyliidae (Diptera: Asiloidea). Bull. Am. Mus. Nat. Hist. 219: 1-191.

Yeates, D.K. 2002. Relationships of extant lower Brachycera (Diptera): a quantitative synthesis of morphological characters. Zool. Scr. 31: 105-121.

Yeates, D.K., B.M. Wiegmann. 1999. Congruence and controversy: toward a higher-level phylogeny of Diptera. Annu. Rev. Entomol. 44: 397-428.

Yoder, A.D., J.A. Irwin, B.A. Payseur. 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. Syst. Biol. 50: 408-424.

Zuker, C., A.F. Cowman, G.M. Rubin. 1985. Isolation and structure of a rhodopsin gene from *D. melanogaster*. Cell 40: 851-858.

Zuker, C.S., C. Montell, K. Jones, T. Laverty, G.M. Rubin. 1987. A rhodopsin gene expressed in photoreceptor cell R7 of the *Drosophila* eye: Homologies with other signal-transducing molecules. J. Neurosci. 7: 1550-1557.

Table 1. Taxa sampled for 28S rDNA, EF-1α, Opsin, and CAD. Grey indicates sequences that were combined for phylogenetic analyses. Check marks indicate presence of sequences without GenBank accession numbers.

	Source		GenBank Accession Numbers			
Taxon	**MEI No.**	**Distribution**	**28S**	**EF-1α**	**Opsin**	**CAD**
Bombyliidae						
Mythicomyia	N/A	California		AY267548	AY267578	AY267613
Scenopinidae						
Stenomphrale teutankhameni	029578	Israel	AF147824	AF148085	AY267587	AY267620
Therevidae						
Phycinae						
Hemigephyra	124388	S. Africa	AY267552	AY267547	AY267573	AY267608
Henicomyia hubbardi	027097	Arizona	AF147828	√		
Henicomyia sp.	139545	Arizona			AY267574	AY267609
Neotabuda	089473	S. Africa	√	AF148095	AY267579	AY267614
Ruppellia multisetosis	011183	Madagascar			AY267585	AY267618
Ruppellia basalis	089476	S. Africa	AF147833			
Ataenogera	N/A		√	√	AY267562	AY267601
Ataenogera abdominalis	124662	Argentina			AY267563	AY267602
Ataenogera abdominalis	N/A	Guatemala	AF147829	AF148089		
Phycus sp.	050693	Israel			AY267583	AY267616
Phycus niger	089478	Namibia		AF148090		
Pherocera	N/A	California			AY267584	AY267617
Pherocera	N/A		√	√		
Agapophytinae sensu lato						
Acraspisa sp. 1	N/A	Australia			AY267556	AY267596
Acraspisa sp.	089456	Australia	AF147837	AF148107		
Agapophytus bicolor	N/A	Australia			AY267559	AY267598
Agapophytus sp.	053710	Australia	AF147840	AF148106		
Actenomeros	N/A	Australia	AY267550	AY267545	AY267557	AY267597
Entesia	106357	Chile	√	AF148099	AY267572	AY267607
Melanothereva	106857	Chile			AY267577	AY267612
Melanothereva	106858	Chile	√			
Chromolepida pruinosa	090267	Guatemala			AY267568	
	090266	Guatemala				AY267605
Chromolepida pruinosa	090268	Guatemala	AF124487	AF124488		
New Genus S	N/A	Australia	AY267553	AY267549	AY267586	AY267619
New Genus M	N/A	Australia	AY267551	AY267546	AY267566	AY267604
Therevinae						
Anabarhynchus sp. 1	N/A	Australia			AY267560	AY267599
Anabarhynchus tristis	053735	Australia	AF147847	AF148108		
Anabarhynchus sp. 2	N/A	New Caledonia	√	√	AY267561	AY267600
Brachylinga	090244	Guatemala			AY267564	AY267603
Brachylinga	090258	Guatemala	AF147849	AF148114		
Cyclotelus pictipennis	089439	Illinois			AY267571	AY267606
Cyclotelus pictipennis	089430	Illinois	AF147852	AF148112		
Penniverpa	090271	Guatemala	AF147851	AF148109	AY267582	AY267615
Irwiniella velutina	011319	Madagascar	√	√	AY267576	AY267611
Hoplosathe frauenfeldi	050690	Israel	AF147854	AF148118	AY267575	AY267610
Thereva sp.	089516	Colorado			AY267589	√
Thereva noblitata	N/A	England	AF147856	AF148121		

Table 2. Primer sequences of opsin, CAD, 28S rDNA, and EF-1α used for PCR and sequencing. Degenerate positions are noted by their IUB single-letter code: R=A/G; Y=C/T; M=A/C; N= A/G/C/T.

Gene	Primer	Sequence (5'→3')
Opsin	ops102f	ATGATHACNAAYACNCCNATGAT
	ops336r	CTDATNCCRTANACDATNGGRTT
	ops1f	ATTTGGTCGATGTGCATGAT
	ops3r	GCAATTATGAACCAATAAG
	lwrhf	AATTGCTATTAYGARACNTGGGT
	M13(-21)-lwrhf	TGTAAAACGACGGCCAGTAATTGCTATTAYGARACNTGGGT
	M13f	TGTAAAACGACGGCCAGT
	lwrhr	ATATGGAGTCCANGCCATRAACCA
	M13-lwrhr	CAGGAAACAGCTATGACCATATGGAGTCCANGCCATRAACCA
	M13r	CAGGAAACAGCTATGACC
CAD	CAD320f	ATHTTYGGNATYTGYYTGGGNCAYCA
	CAD329f	ATHGGNTGYAARACNTAYAARATG
	CAD338f	ATGAARTAYGGYAATCGTGGHCAYAA
	CAD581 repl.f	GGWGGWCAAACWGCWYTMAAYTGYGG
	CAD654r	TCYTTCCANCCYTTYARSGATTTRTC
	CAD680r	AANGCRTCNCGNACMACYTCRTAYTC
	CAD691r	GGRTCRAARTTYTCCATRTTRCA
	CAD835r	CATNACYTCNCCNACRCTYTTCAT
	CAD843r	GCYTTYTGRAANGCYTCYTCRAA
28S rDNA	rc28H	CTACTATCCAGCGAAACC
	rc28Q	GGACATTGCCAGGTAGGGAGTT
	28K	CTTCGATGTCGGCTCTTG
	28Z	GCAAAGGATAAGCTTCAGTGG
EF-1α	EF-1α 3' PCR	ACAGCVACKGTYTGYCTCATRTC
	EF-1α 5' PCR	GARCGTGGTATCACCATYGAYAT
	Joe-2	CCGTGGTWCAAGGGATGG
	Shemp	TCCRATACCNCCRATTTTGTA
	EF2	GGATGGCAYGGYGACAACATG
	EF5	CTCATATCACGTACAGCRAARCG
	EF4	GARCGTGGTATYACMATTGA
	EF6	CWCCAGTTTCWACACGWCC
	EF51	CATGTTGTCRCCRTGCCATCC

Table 3. Descriptive statistics for phylogenetic analyses.

Taxon Analysis Set	28S rDNA + EF-1α	28S rDNA + EF-1α (no third pos.)	Opsin + CAD	Opsin + CAD (no third pos.)	28S rDNA, EF-1α, Opsin, CAD	28S rDNA, EF-1α, Opsin, CAD (no third pos.)
Length (bp)	2258	1927	1821	1214	4079	3141
No. variable sites (% of total)	566 (25)	276 (14)	959 (53)	374 (31)	1525 (37)	650 (21)
No. parsimony informative sites (% of total)	377 (17)	134 (7)	767 (42)	203 (17)	1144 (28)	337 (11)
%A	29.9	30.2	30.8	30.7	30.3	30.4
%C	18.9	19.1	20.1	20.6	19.5	19.7
%G	22.1	24.2	21.7	23.5	22.0	23.9
%T	29.2	26.5	27.4	25.2	28.3	26.0

Table 4. Tree statistics for unweighted maximum parsimony trees (MPTs).

Taxon Analysis Set	28S rDNA+ EF-1α	28S rDNA +EF-1α (no third pos.)	Opsin+ CAD	Opsin+ CAD (no third pos.)	28S rDNA, EF-1α , Opsin, CAD	28S rDNA, EF-1α, Opsin, CAD (no third pos.)
MPT length	1640	500	4504	920	6199	1435
No. MPTs	12	145	9	15	1	12
CI	0.469	0.658	0.366	0.528	0.390	0.568
RI	0.483	0.616	0.383	0.471	0.401	0.510

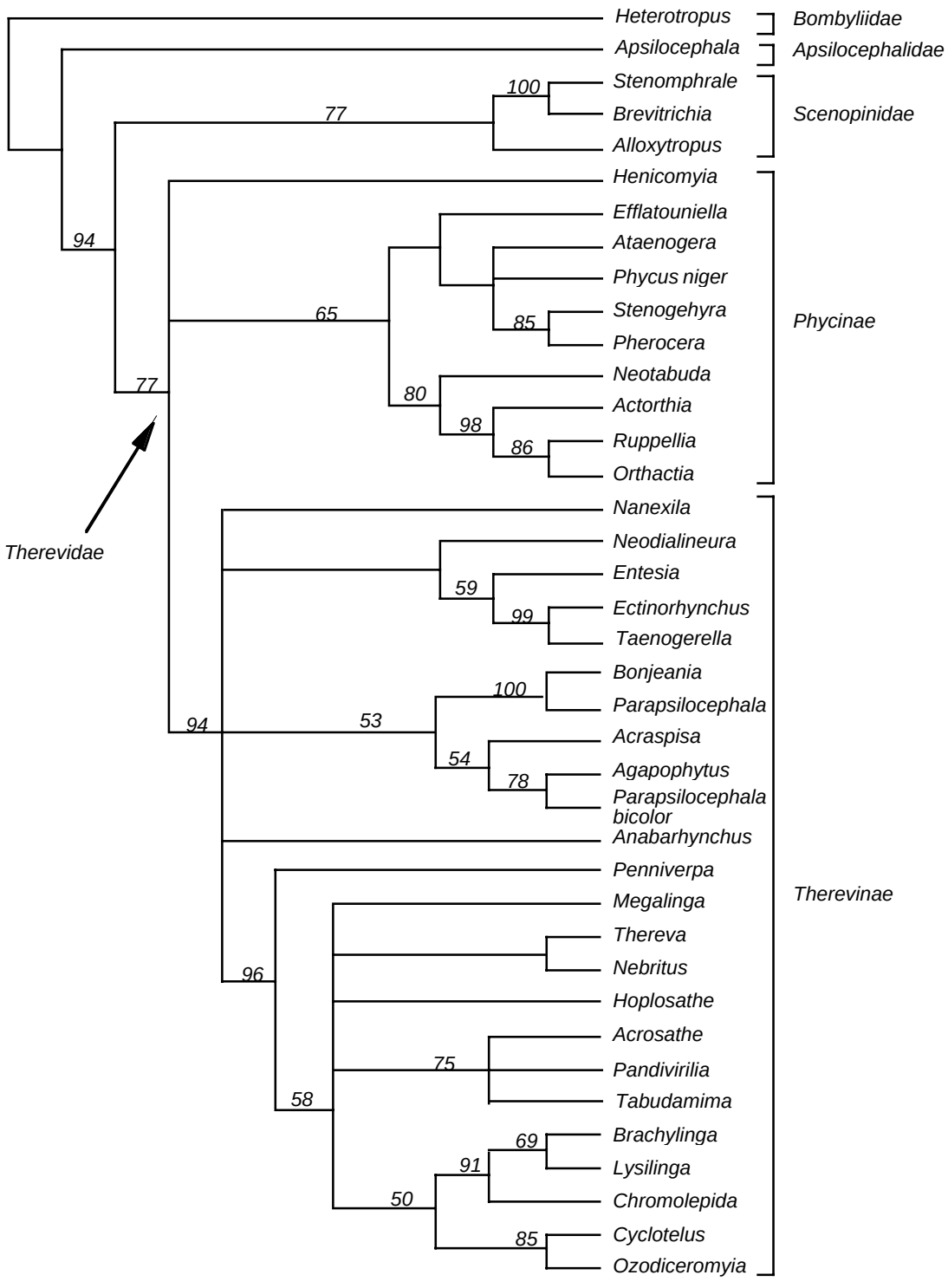

Figure 1. Phylogenetic relationships based on combined 28S rDNA and EF-1α gene sequences of 38 genera from study conducted by Yang et al. (2000). Strict consensus of 14 EPTs with the bootstrap values (200 replicates) shown above each branch.

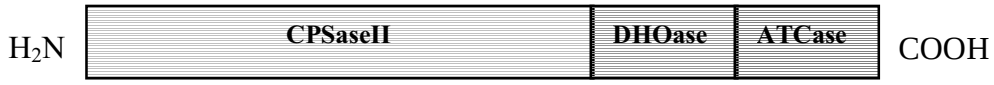

Figure 2. CAD domain organization.

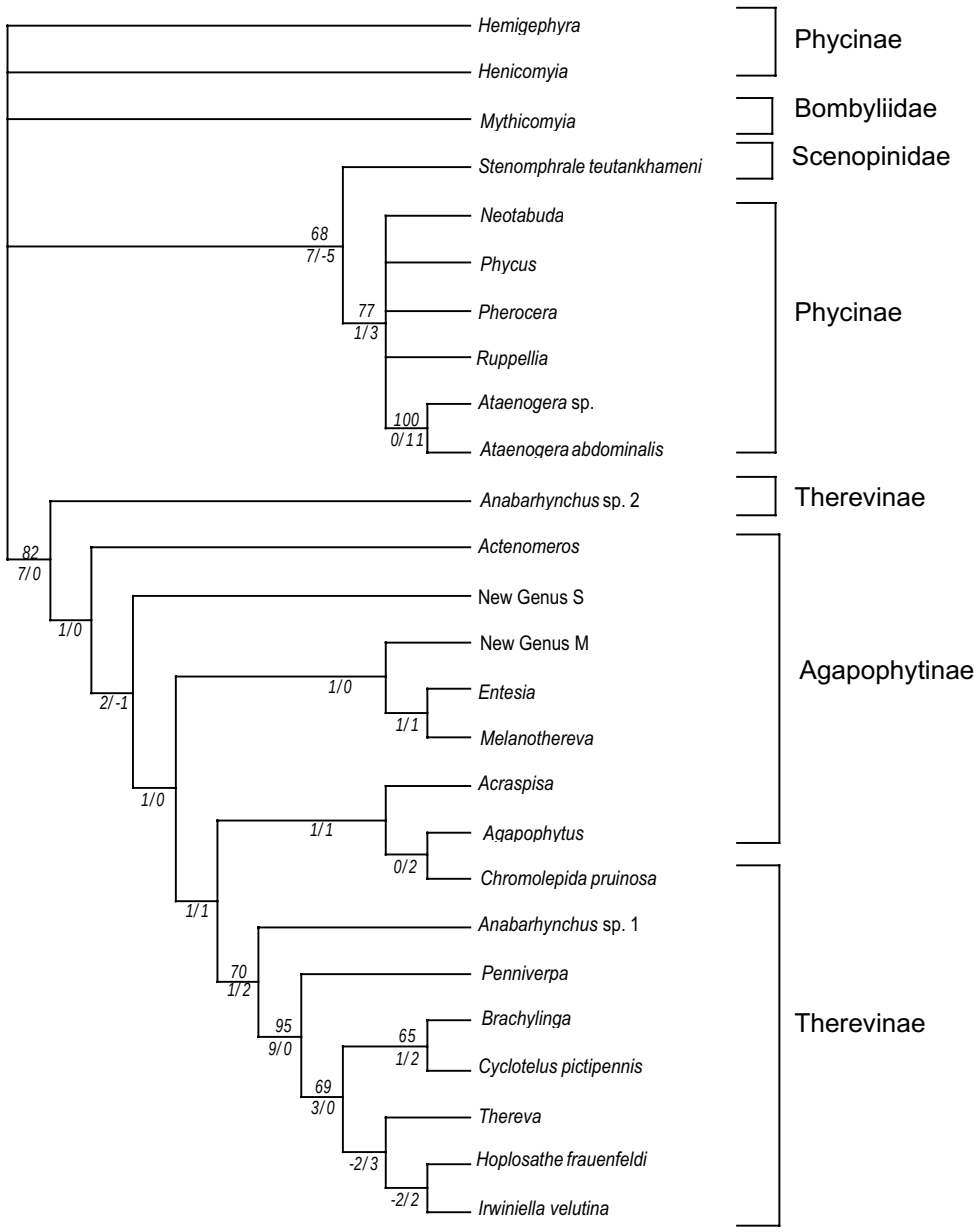

Figure 3. Phylogenetic relationships based on combined analysis of 28S rDNA and EF-1α gene sequences of 26 taxa in 24 genera. The tree was estimated using unweighted parsimony analysis of nucleotide sequences, implemented in PAUP*, version 4.0. Strict consensus of 12 EPTs; tree length = 1640; CI = 0.469; RI = 0.483. Bootstrap percentages (calculated with 1000 replicate heuristic searches) ≥ 50% are shown above each branch and Bremer support values are below (EF-1α/ 28S).

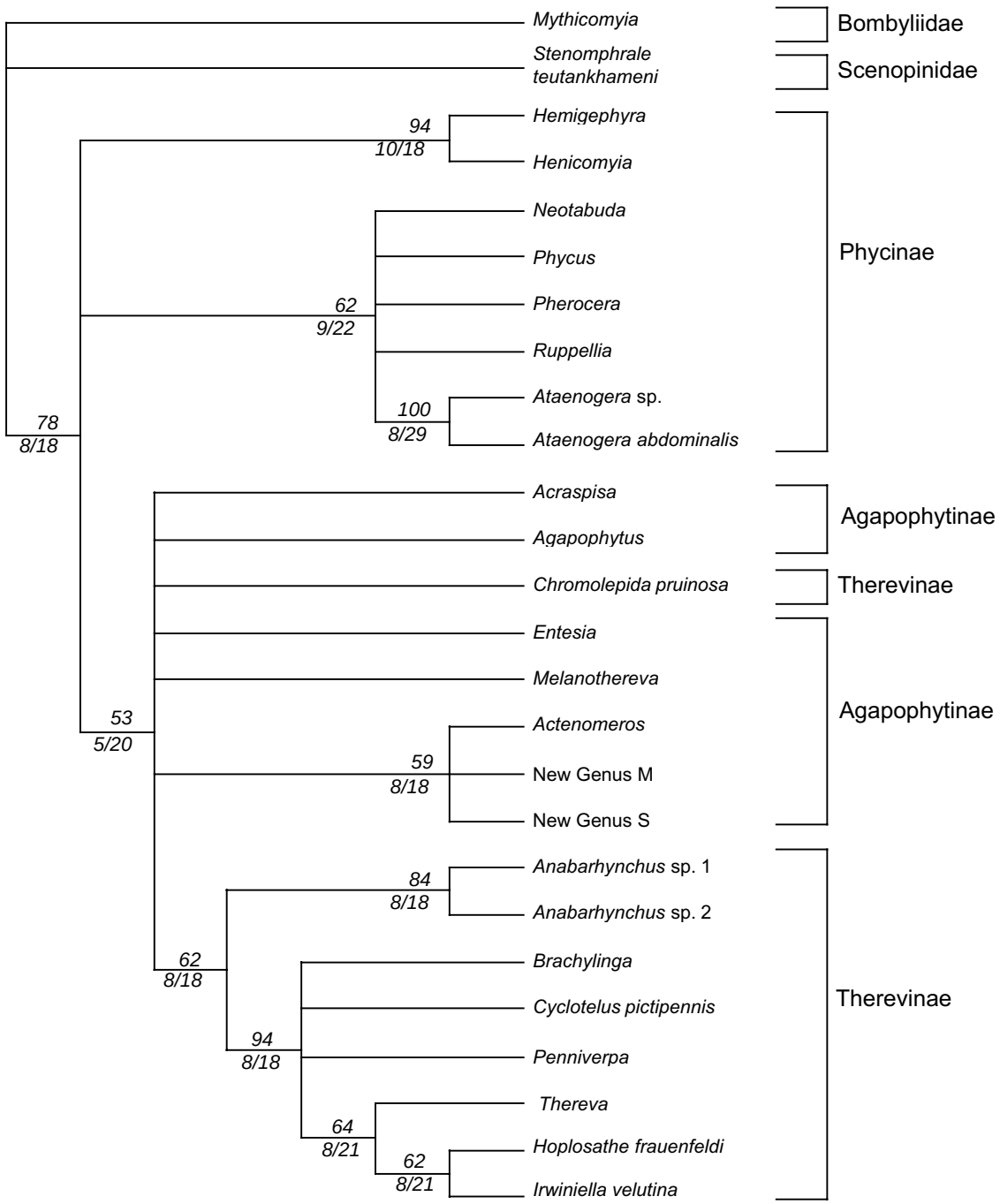

Figure 4. Phylogenetic relationships based on combined analysis of 28S rDNA and EF-1α gene sequences of 26 taxa in 24 genera, excluding third nucleotide positions. The tree was estimated using unweighted parsimony analysis of nucleotide sequences, implemented in PAUP*, version 4.0. Strict consensus of 145 EPTs; tree length = 500; CI = 0.658; RI = 0.616. Bootstrap percentages (calculated with 1000 replicate heuristic searches) ≥ 50% are shown above each branch and Bremer support values are below (EF-1α/ 28S).

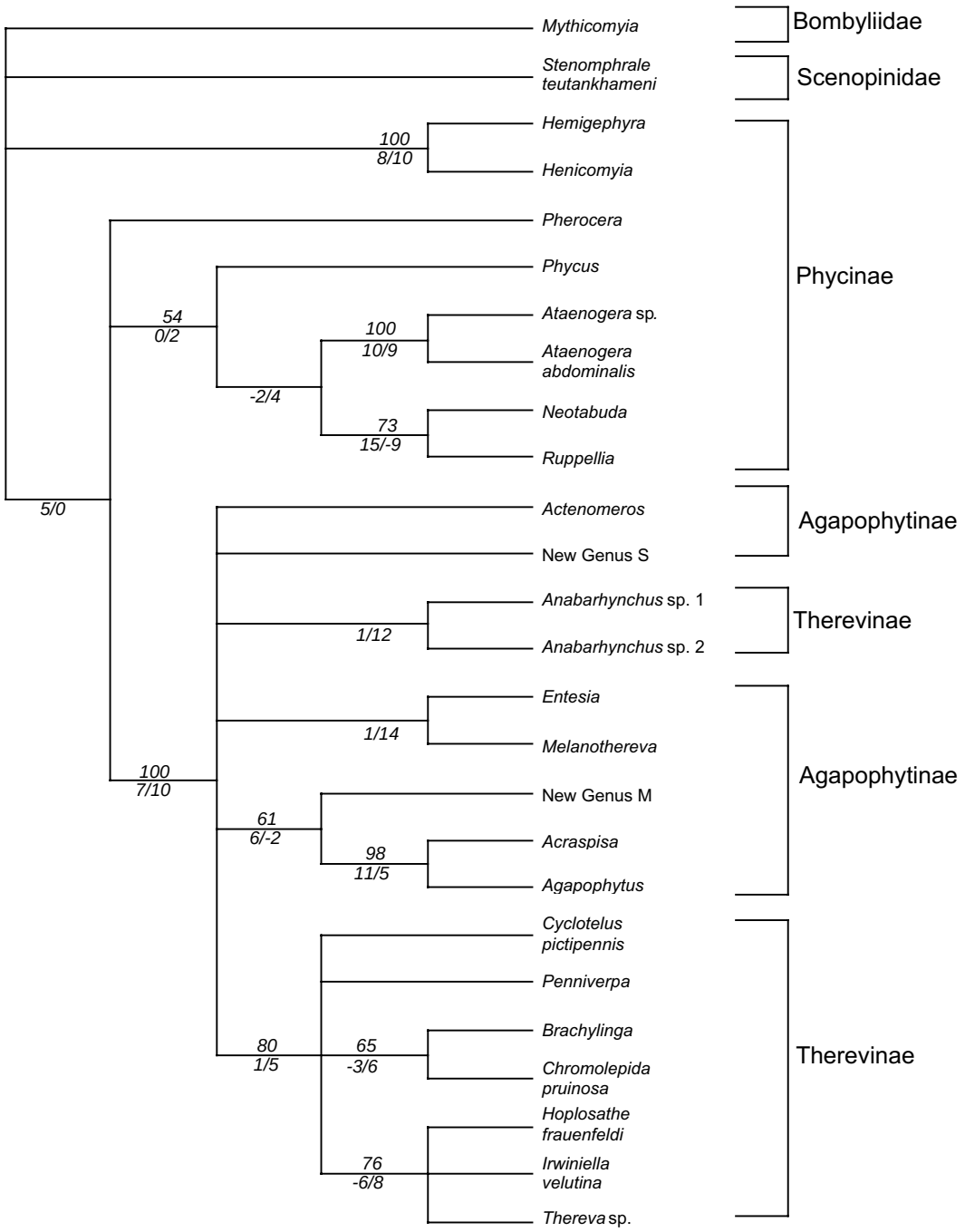

Figure 5. Phylogenetic relationships based on combined analysis of opsin and CAD gene sequences of 26 taxa in 24 genera. The tree was estimated using unweighted parsimony analysis of nucleotide sequences, implemented in PAUP*, version 4.0. Strict consensus of 9 EPTs; tree length = 4504; CI = 0.366; RI = 0.383. Bootstrap percentages (calculated with 1000 replicate heuristic searches) ≥ 50% are shown above each branch and Bremer support values are below (opsin/CAD).

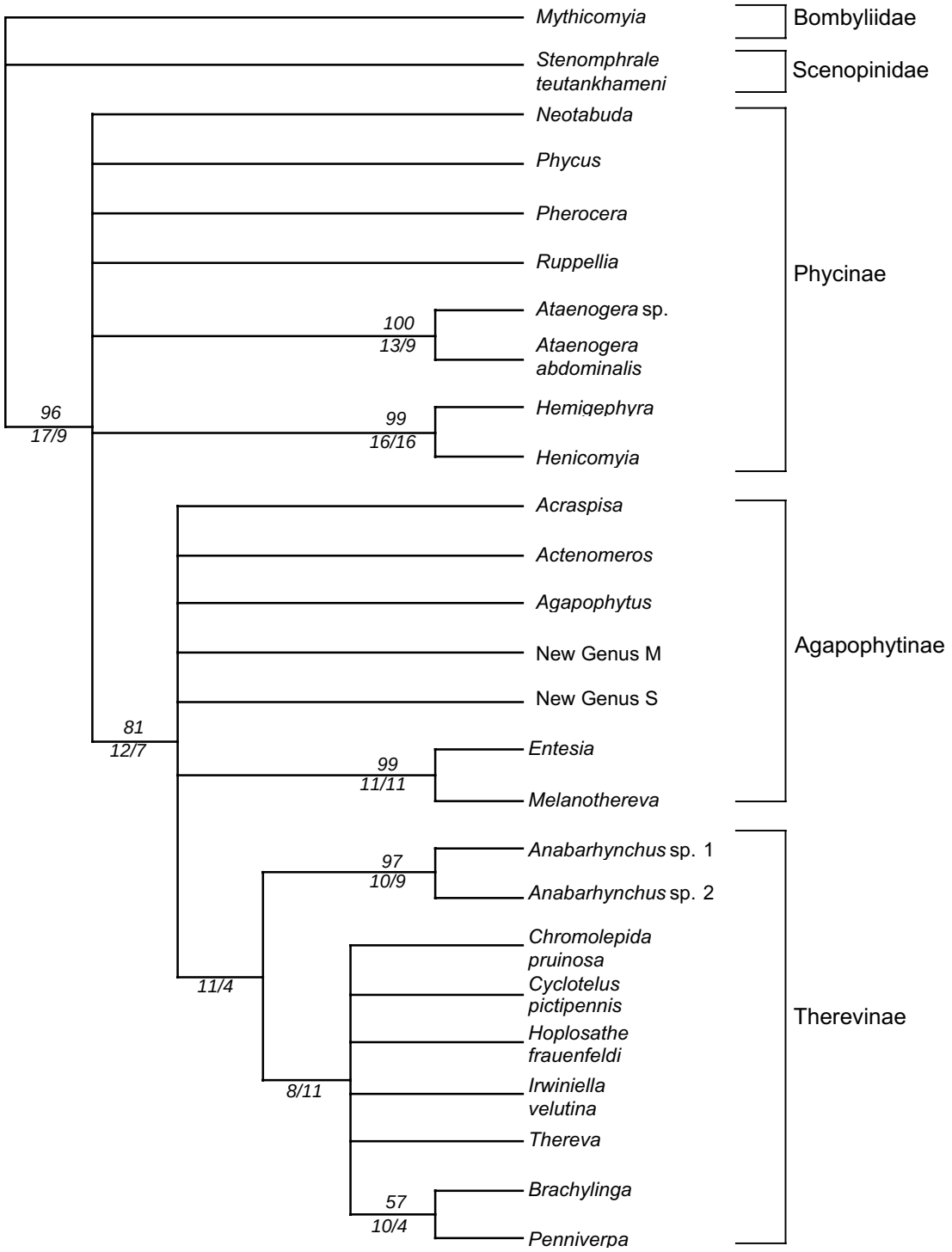

Figure 6. Phylogenetic relationships based on combined analysis of opsin and CAD gene sequences of 26 taxa in 24 genera, excluding third nucleotide positions. The tree was estimated using unweighted parsimony analysis of nucleotide sequences, implemented in PAUP*, version 4.0. Strict consensus of 15 EPTs; tree length = 920; CI = 0.528; RI = 0.471. Bootstrap percentages (calculated with 1000 replicate heuristic searches) ≥ 50% are shown above each branch and Bremer support values are below (opsin/CAD).

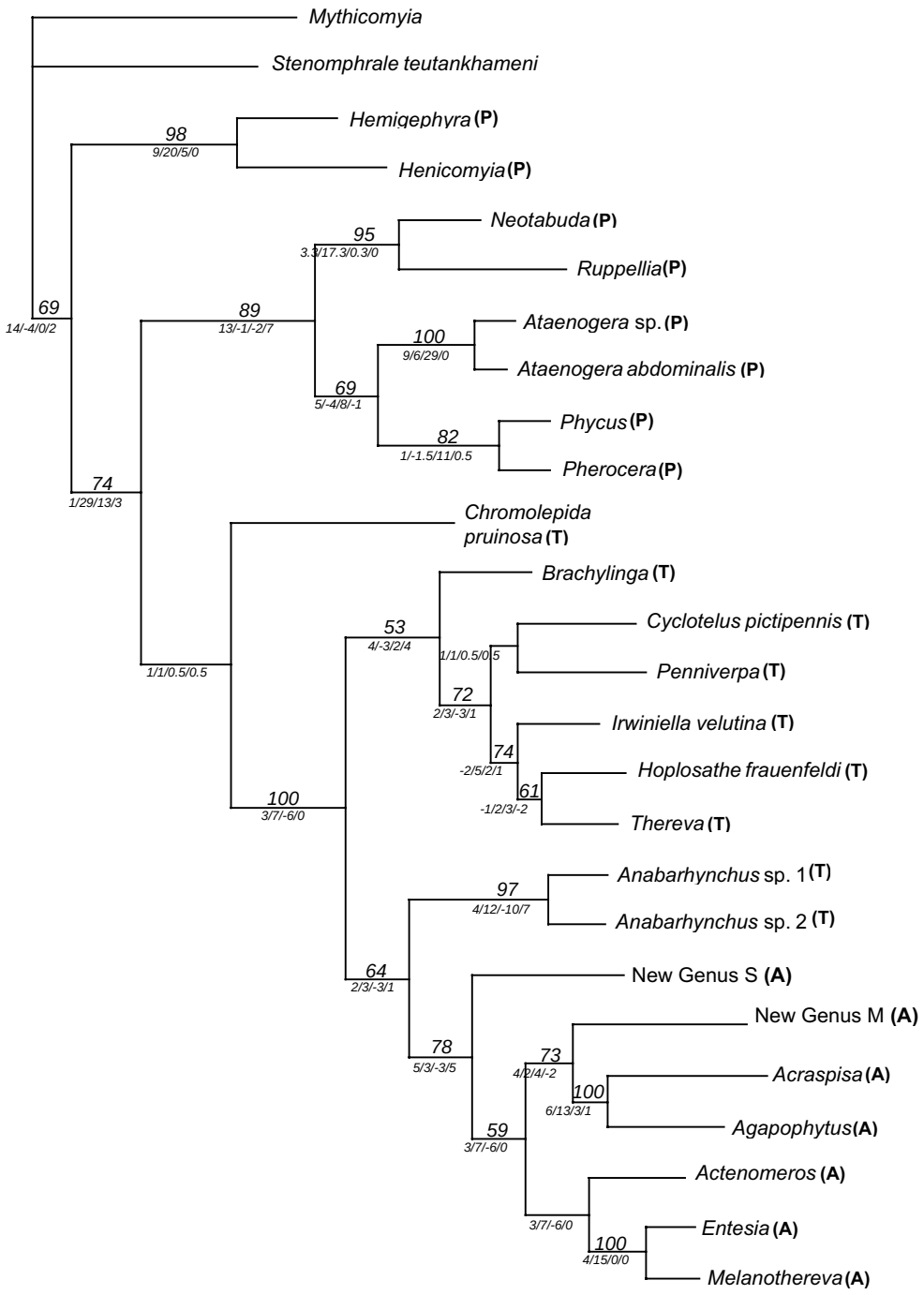

Figure 7. Phylogenetic relationships based on combined analysis of 28S rDNA, EF-1α, opsin and CAD gene sequences of 26 taxa in 24 genera. Subfamilies denoted as follows: Phycinae = P; Therevinae = T; Agapophytinae = A. The tree was estimated using unweighted parsimony analysis of nucleotide sequences, implemented in PAUP*, version 4.0. Single MPT; tree length = 6199; CI = 0.390; RI = 0.401. Bootstrap percentages (calculated with 1000 replicate heuristic searches) ≥ 50% are shown above each branch and Bremer support values are below (opsin/ CAD/ EF-1α/ 28S).

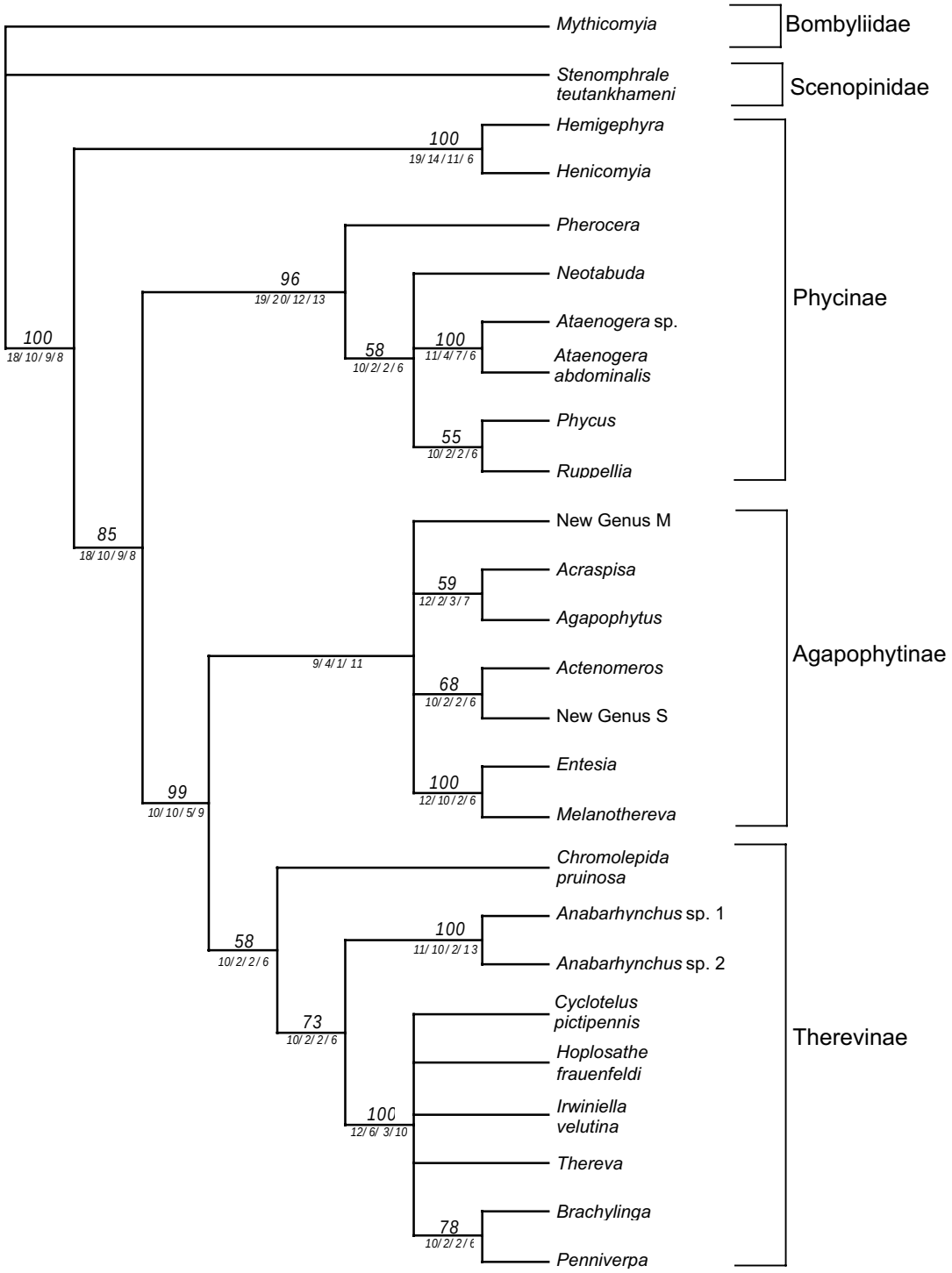

Figure 8. Phylogenetic relationships based on combined analysis of 28S rDNA, EF-1α, opsin and CAD gene sequences of 26 taxa in 24 genera, excluding third nucleotide positions. The tree was estimated using unweighted parsimony analysis of nucleotide sequences, implemented in PAUP*, version 4.0. Strict consensus of 12 EPTs; tree length = 1435; CI = 0.568; RI = 0.510. Bootstrap percentages (calculated with 1000 replicate heuristic searches) ≥ 50% are shown above each branch and Bremer support values are below.

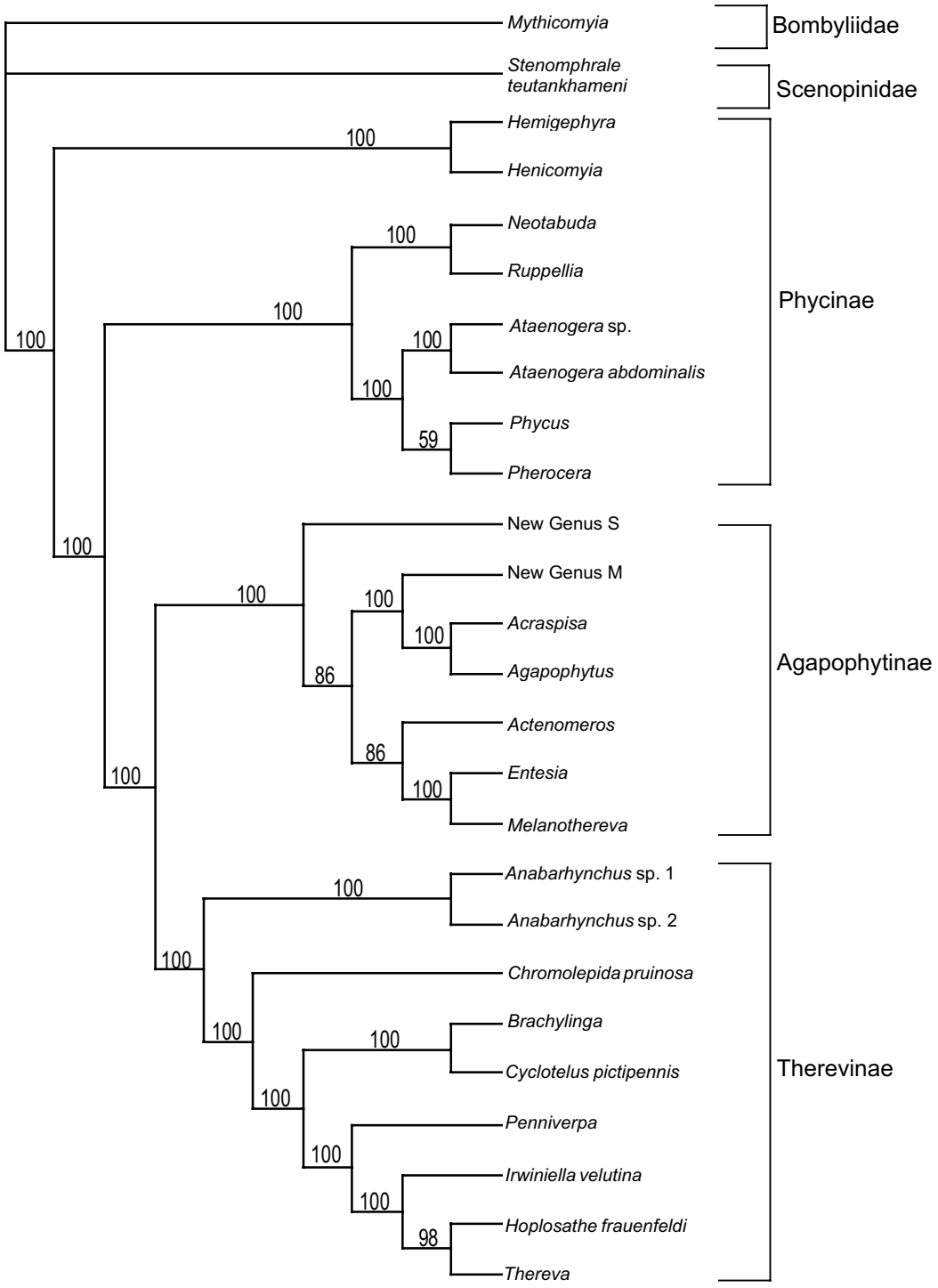

Figure 9. The 50% majority rule consensus tree from the combined analysis of 28S rDNA, EF-1α, opsin, and CAD nucleotide gene sequences of 26 taxa in 24 genera. Numbers on nodes represent posterior probability values.