Molecular Phylogenetics and Evolution
Vol. 12, No. 2, July, pp. 168–176, 1999
Article ID mpev.1998.0606, available online at http://www.idealibrary.com on IDEAL®

The Major Opsin in Bees (Insecta: Hymenoptera): A Promising Nuclear Gene for Higher Level Phylogenetics

Patrick Mardulyn*,[1] and Sydney A. Cameron*,†,[2]

*Department of Entomology and †Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas 72701

Received July 9, 1998; revised September 18, 1998

We report the phylogenetic utility of the nuclear gene encoding the long-wavelength opsin (LW *Rh*) for tribes of bees. Aligned nucleotide sequences were examined in multiple taxa from the four tribes comprising the corbiculate bees within the subfamily Apinae. Phylogenetic analyses of sequence variation in a 502-bp fragment (approx 40% of the coding region) strongly supported the monophyly of each of the four tribes, which are well established from previous studies of morphology and DNA. Trees estimated from parsimony and maximum likelihood analyses of LW *Rh* sequences show a strongly supported relationship between the tribes Meliponini and Bombini, a relationship that has been found uniformly in studies of other genes (28S, 16S, and cytochrome *b*). All of the tribal clades as well as relationships among the tribes are supported by high bootstrap values, suggesting the utility of LW *Rh* in estimating tribal and subfamily rank for these bees. The sequences exhibit minimal base composition bias. Both 1st + 2nd and 3rd position sites provide information for estimating a reliable tree topology. These results suggest that LW *Rh*, which has not been reported previously in studies of organismal phylogenetics, could provide important new data from the nuclear genome for phylogeny reconstruction. © 1999 Academic Press

Key Words: **major opsin; long wavelength rhodopsin; LW *Rh*; nuclear gene; bees; Hymenoptera; molecular phylogeny.**

INTRODUCTION

There is a great need for comparative data from the nuclear genome to test for concordance with data gathered from organellar genomes and from morphology. Yet, other than the widely used nuclear ribosomal DNA (rDNA) subunits, only a handful of nuclear genes have been tested for phylogenetic utility above the level of populations (Thomas and Hunt, 1993 for A*dh;* Cho *et al.,* 1995; Mitchell *et al.,* 1997; Danforth and Ji, 1998 for EF-1α; Friedlander *et al.,* 1996 for PEPCK; Fang *et al.,* 1997; Friedlander *et al.,* 1998 for DDC; and Gupta, 1995; Waters, 1995 for some of the heat shock proteins; see review in Brower and DeSalle, 1994).

In this report, we assess the phylogenetic utility of one of the family of visual pigment genes (opsins or rhodopsins) found in honey bees (*Apis mellifera*) that encodes the long-wavelength visual pigment, LW *Rh,* also known as the green or major opsin (Chang *et al.,* 1996). The LW *Rh* gene has been identified recently in *Apis* (Chang *et al.,* 1996) and in *Mantis* (*Sphodromantis* spp.) (Towner and Gärtner, 1994). Two other genes encoding visual pigments have been characterized in *Apis,* a blue-sensitive opsin and a UV-sensitive opsin (Townson *et al.,* 1998). All three *Apis* visual pigment genes encode light-absorbing proteins that are maximally sensitive to light within the range of expected wavelengths: LW *Rh* at 540 nm, blue at 439 nm, and UV at 353 nm (Townson *et al.,* 1998). In concert, the opsins govern an organism's visual sensitivity. To date, the phylogenetic interest in visual pigment genes has focused primarily on the molecular evolution of the different members of the multigene family (Chang *et al.,* 1995; Crandall and Cronin, 1997; Yokoyama, 1995). To our knowledge, there are no published reports involving the use of any of the opsin (or rhodopsin) genes for phylogenetic analysis of organisms.

Herein, we examine the utility of the putative LW *Rh* gene for the novel use of reconstructing phylogenetic relationships of organisms. Our study organisms for this gene are the bees belonging to the monophyletic clade known as the corbiculate bees, those with a specialized pollen-carrying structure called a corbicula. The corbiculate clade falls within the subfamily Apinae (Hymenoptera: Apidae) and contains four tribes, including the highly social honey bees and stingless bees, which exhibit some of the most elaborate social behavior known in insects (Michener, 1974). The monophyly of each tribe is strongly supported by morphological characters (Michener, 1990) and DNA sequence data

[1] Present address: Department of Biology, Free University of Brussels, B-1050 Brussels, Belgium.
[2] To whom correspondence should be addressed. Fax: 501-575-2452. E-mail: scameron@comp.uark.edu.

1055-7903/99 $30.00

(Cameron, 1992, 1993; Koulianos *et al.*, 1999). Similar patterns of relationship among the tribes have been estimated independently from two mitochondrial genes (16S rDNA, Cameron, 1993; cytochrome *b*, Koulianos *et al.*, 1999) and one nuclear gene (28S rDNA, Sheppard and McPheron, 1991). All three genes strongly support a clade (Meliponini + Bombini). Two of them unambiguously support the grouping of this clade with the Euglossini, while the third gene, cytochrome *b*, supports the same grouping or an alternative clade (Euglossini + Apini), depending on whether the amino acid or the nucleotide sequences are analyzed (Koulianos *et al.*, 1999). The topological congruence of the tribal tree topologies from multiple sources of data enables us to test for concordance of the LW *Rh* corbiculate bee trees against a well corroborated molecular phylogeny. We first characterize the nucleotide substitution patterns of a fragment of the LW *Rh* gene and then test its phylogenetic utility by examining concordance of the LW *Rh* trees with those estimated from the other nuclear and mitochondrial genes.

MATERIALS AND METHODS

Bees Examined

Opsin sequences were obtained from multiple exemplars representing each of the four tribes of corbiculate bees. This group includes the highly eusocial honey bees (Apini, 1 genus, 6 species; Alexander, 1991) and stingless bees (Meliponini, 21 genera, 300–430 species; Michener, 1990 [Meliponini has also been divided into as many as 55 genera; Camargo and Pedro, 1992]), the intermediately social bumble bees (Bombini, 1 genus, 239 species; Williams, 1998), and the solitary orchid bees (Euglossini, 5 genera, 174 species; Kimsey, 1987). Together, they form a monophyletic group within the subfamily Apinae (family Apidae) (Roig-Alsina and Michener, 1993). The use of exemplars to represent the corbiculate tribes is justified on the basis that each tribe has been recognized as a monophyletic group by several independent studies of morphology and DNA (summarized in Michener, 1990, and references therein; Cameron, 1993). However, taxon sampling is a critical aspect in phylogenetic studies. Hence, we have sampled across a greater diversity of tribal genera and species than in prior molecular or recent morphological analyses of the corbiculate bees. For the Apini (a monogeneric tribe) we have sampled from 3 of the 6 described species. For Euglossini we have sampled from 4 of the 5 genera. For Bombini (another monogeneric tribe) we have sampled from 3 of the 35 described subgenera. For Meliponini we have sampled from 4 of the 21 genera. Table 1 indicates the taxonomic affinity, collection site, collector, and GenBank accession number for each species examined. Two outgroups were selected from other monophyletic tribes within the Apidae (Roig-Alsina and Michener, 1993): *Melissodes* (Eucerini: Apinae) and *Xylocopa* (Xylocopini: Xylocopinae).

Voucher specimens for all taxa used in this investiga-

TABLE 1

Taxa Examined

Subfamily/tribe	Species	Collection site	Collector	GenBank Accession No.
Apinae				
Apini (6)	*Apis mellifera*	Arkansas, USA	SAC	AF091732
	Apis nigrocincta	Sulawesi	GWO	AF091728
	Apis dorsata	India	SAC	AF091733
Bombini (239)	*Bombus pennsylvanicus*	Arkansas, USA	SAC	AF091727
	Bombus avinoviellus	India	SAC	AF091719
	Bombus terrestris	Great Britain	HS	AF091722
Meliponini (300–430)	*Trigona hypogea*	Brasil	SAC	AF091724
	Scaptotrigona depilis	Brasil	SAC	AF091729
	Tetragona dorsalis	Panama	DWR	AF091726
	Lestrimelitta limao	Panama	DWR	AF091723
Euglossini (174)	*Eufriesea caerulescens*	Mexico	SAC	AF091725
	Euglossa imperialis	Panama	RSH	AF091720
	Exaerete frontalis	Panama	RSH	AF091718
	Eulaema meriana	Cost Rica	SAC	AF091721
Outgroups				
Apinae				
Eucerini	*Melissodes rustica*	Arkansas, USA	SAC	AF091731
Xylocopinae				
Xylocopini	*Xylocopa virginica*	Missouri, USA	SAC	AF091730

Note. Numbers in parentheses in the left column, next to the four corbiculate bee tribes, are approximate numbers of species in those tribes. Collector abbreviations: SAC, Sydney Cameron; GWO, Gard Otis; DWR, David Roubik, RSH, Regula Schmid-Hempel, HS, Horst Schwarz.

tion are deposited in the Arthropod Museum at the University of Arkansas, Fayetteville.

PCR and DNA Sequencing

Genomic DNA was extracted from fresh, frozen (−80°C), and ethanol-preserved tissue from each insect's thorax, abdomen, or legs. Tissue was ground in an SDS homogenization buffer, incubated for 1–2 h with proteinase K at 60°C, followed by four phenol/chloroform extractions, ethanol precipitation, and resuspension in TE buffer (10 mM Tris, 1 mM EDTA). A ±700-bp fragment of the opsin gene was PCR amplified for each species using primers LWRhF-5′ AAT TGC TAT TAY GAR ACN TGG GT 3′ and LWRhR-5′ ATA TGG AGT CCA NGC CAT RAA CCA 3′ (developed by Donat Agosti, AMNH), which were based on published opsin sequences of *Drosophila pseudoobscura* (Carulli and Hartl, 1992). These primers amplified a fragment apparently corresponding to nucleotide positions 873–1825, as numbered in the *D. pseudoobscura Rh1* sequence (Carulli and Hartl, 1992), and found later to match positions 421–922 of the recently published *A. mellifera* long-wavelength opsin sequence when the introns are excluded (Chang *et al.*, 1996). PCR conditions followed an initial denaturation step of 30 s at 94°C, 35 cycles of a 60-s denaturation at 94°C, 60-s annealing at 60°C, and 60-s extension at 72°C, with a final extension step of 2 min at 72°C. PCR products were purified using the Wizard PCR Preps DNA Purification System (Promega). Sequencing was conducted with an ABI 377 automated sequencer, using the PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit FS (Perkin–Elmer) according to manufacturer's specifications. Both strands were sequenced for all taxa using the PCR primers LWRhF and LWRhR.

Sequence Variability and Phylogenetic Analyses

Introns were identified by comparing the apine bee sequences to known opsin sequences (Chang, 1996) and were excluded from the data set prior to phylogenetic analyses. Nucleotide sequences were aligned easily by eye with no inferred gaps. MacClade 3.07 (Maddison and Maddison, 1992) was used to estimate the frequency distribution of the observed number of substitutional changes per character at first plus second (1st + 2nd) and at third (3rd) codon positions. Base composition frequencies for each codon position and uncorrected pairwise sequence divergences were calculated with a test version of PAUP*4 (test versions 4.0D63-64, provided by D. L. Swofford). A χ^2 test of homogeneity of base frequencies across taxa was also performed using PAUP*4.

All phylogenetic analyses were implemented in PAUP*4. Unless otherwise indicated, all parsimony analyses were run with the branch-and-bound algorithm to ensure that all most-parsimonious trees were found. An analysis using the maximum likelihood (ML) method is described in detail below. All parameters used in the ML model were estimated using the likelihood criterion. Bootstrap values (BV) (Felsenstein, 1985) were estimated to provide measures of relative support for each clade. For parsimony analyses, these were based on 400 replicate branch-and-bound searches. ML analyses employed heuristic searches with as-is addition sequence and TBR swapping.

In all organisms studied thus far, the genes encoding visual pigments occur as members of a multigene family; hence, there is the potential for sequencing nonorthologous (paralogous) copies in the different taxa examined (Yokoyama, 1995; Chang *et al.*, 1996; Crandall and Cronin, 1997; Townson *et al.*, 1998). To examine this possibility and to verify the affinity of the bee sequences to other known long-wavelength sequences, we retrieved additional amino acid sequences from GenBank representing different copies of visual pigment genes for a diversity of insects (Table 2). These were aligned with a subset of our bee sequences (one species from each apine tribe and one outgroup species) using Clustal W 1.7 (Thompson *et al.*, 1994) and subjected to unweighted parsimony analysis (heuristic search, 100 random addition sequences, TBR swapping). Because our bee sequences represent a fragment of the entire LW *Rh* gene, more than 50% of the characters were missing for those sequences and were coded as such in the resulting data matrix.

Tribal relationships of the corbiculate bees were estimated from parsimony analyses of the entire nucleotide data set as well as the (1st + 2nd) or 3rd positions separately. Because nucleotide composition of most 3rd-position sites does not influence the amino acid composition, their evolution is less constrained. To explore the difference in phylogenetic signal between (1st + 2nd) and 3rd positions, an incongruence length difference (ILD) test (Farris *et al.*, 1994) was performed between those two classes of sites (1000 replications), using PAUP*4. Differential weighting of transitions (TI) and transversions (TV) may be justified if saturation by substitutional type occurs (Milinkovitch *et al.*, 1996; Swofford *et al.*, 1996). To examine whether TI or TV were saturated, we plotted the number of TI against the number of TV for all possible pairs of taxa, for each codon position. A weighted parsimony analysis was performed using the TI/TV plot as a guide to define a weighting strategy (see Results).

Because parsimony methods have been shown to be inconsistent under certain conditions (Felsenstein, 1978; Hendy and Penny, 1989; Kim, 1996; Huelsenbeck, 1997), we also performed a maximum likelihood analysis, which incorporates an empirically derived model of substitutional change and takes branch lengths into account. A preliminary ML search was implemented with a HKY85+Γ_4 model of DNA substitution (Hasegawa *et al.*, 1985; Yang, 1993), which corrects for multiple substitutions and allows for different base

TABLE 2

Sequences from Different Opsin Copies of Insects Retrieved from GenBank

Species	Insect order	Sequence label	Accession Number	Reference
Drosophila melanogaster	Diptera	D. mel Rh1	K02315	O'Tousa *et al.*, 1985
D. melanogaster	Diptera	D. mel Rh2	M12896	Cowman *et al.*, 1986
D. melanogaster	Diptera	D. mel Rh3	M17718	Zuker *et al.*, 1987
D. melanogaster	Diptera	D. mel Rh4	M17719	Montell *et al.*, 1987
D. melanogaster	Diptera	D. mel Rh5	U67905	Chou *et al.*, 1996
D. pseudoobscura	Diptera	D. pse Rh1	X65877	Carulli and Hartl, 1992
D. pseudoobscura	Diptera	D. pse Rh2	X65878	Carulli and Hartl, 1992
D. pseudoobscura	Diptera	D. pse Rh3	X65879	Carulli and Hartl, 1992
D. pseudoobscura	Diptera	D. pse Rh4	X65880	Carulli and Hartl, 1992
Apis mellifera	Hymenoptera	LW (Green)	U26026	Chang *et al.*, 1996
Apis mellifera	Hymenoptera	UV	AF004169	Townson *et al.*, 1998
Apis mellifera	Hymenoptera	blue	AF004168	Townson *et al.*, 1998
Camponotus abdominalis	Hymenoptera	ant 1 opsin 1	U32502	Popp *et al.*, 1996
Cataglyphis bombycinus	Hymenoptera	ant 2 opsin 1	U32501	Popp *et al.*, 1996
Camponotus abdominalis	Hymenoptera	ant 1 opsin 2	AF042788	Smith *et al.*, unpublished
Cataglyphis bombycinus	Hymenoptera	ant 2 opsin 2	AF042787	Smith *et al.*, unpublished
Manduca sexta	Lepidoptera	moth	S83264	Chase *et al.*, 1996
Schistocerca gregaria	Orthoptera	Locust	X80072	Towner *et al.*, unpublished
Sphodromantis sp.	Mantodea	mantid	X71665	Towner and Gartner, 1994
Calliphora vicina	Diptera	Calliphora vicina	M58334	Huber *et al.*, 1990

frequencies, TI/TV bias, and among-site rate variation. In the preliminary analysis, we employed a simple heuristic search (as-is addition sequence and TBR swapping) to estimate the TI/TV ratio, shape parameter of the gamma distribution, and proportion of invariable sites from the data. Setting these parameters to their estimated value, a second more thorough ML search was performed with 100 random addition sequences and TBR swapping.

The nucleotide data were also explored using Relative Apparent Synapomorphy Analysis (RASA; Lyons-Weiler *et al.*, 1996) to detect the possible presence of a long-branch attraction problem (Felsenstein, 1978; Hendy and Penny, 1989; Huelsenbeck, 1997) and hence the potential to converge to an incorrect estimate of relationship under some phylogenetic methods such as parsimony. RASA is a regression-based statistical method that tests for the presence of phylogenetic signal in a data set by inferring a slope that depicts the covariation between the apparent cladistic similarity and the phenetic similarity observed between pairs of taxa. The inferred slope is compared with a null slope expected under an equiprobable model (Student's t distribution). This method has the advantage of being tree independent, thus avoiding potential circular reasoning (a tree-based statistic may be misleading if the estimated tree used to calculate that statistic is very different from the "true" phylogeny) (Lyons-Weiler, 1998). We apply an extension of this method, which has proven powerful in detecting long-branch taxa in a data set. Taxa with long branches impart a distinct pattern in the distribution of character states among taxa. This pattern can be observed in the structure of the RASA regression (Lyons-Weiler and Hoelzer, 1997).

RESULTS

Characterization of Nucleotide Sequences

Two introns were present in our amplified opsin fragment. The first of the two introns corresponded to one found in the same relative position in *D. pseudoobscura* (Carulli and Hartl, 1992) and varies in length from 87 to 110 nucleotides. The second was located 114 nucleotides downstream of that found in *D. pseudoobscura* and varies in length from 114 to 189 bp. It was not possible to compare the position of these introns to those of the honey bee sequence of Chang *et al.* (1996) because only the coding region was sequenced in that work. After removal of the introns, the corbiculate bee sequences were 502 bp in length; 132 sites were parsimony informative.

There is little overall base composition bias, although the pattern varies somewhat for each codon position (Table 3). The χ^2 test of homogeneity of base frequencies across taxa resulted in P values of 1.00, 1.00, and 0.12,

TABLE 3

Mean Nucleotide Frequencies per Codon Position (Range Shown in Parentheses)

	A	C	G	T
1st positions	32.6 (31.0–34.1)	13.7 (12.6–16.1)	25.7 (24.6–28.3)	28.0 (26.1–29.3)
2nd positions	22.8 (20.9–24.1)	21.6 (20.3–22.7)	19.4 (18.0–21.0)	36.2 (33.7–38.0)
3rd positions	21.2 (16.2–24.1)	30.3 (27.5–40.6)	22.2 (15.8–27.4)	26.4 (17.5–37.5)

TABLE 4

Uncorrected Pairwise Divergences between Nucleotide Sequences (below Diagonal) and between 3rd Position Sites Only (above Diagonal)

	Eufriesea	*Euglossa*	*Eulaema*	*Exaerete*	*Bombus avi.*	*Bombus pen.*	*Bombus ter.*	*Lestrimel.*	*Scaptotrig.*	*Tetragona*	*Trigona*	*Apis mel.*	*Apis nig.*	*Apis dor.*	*Xylocopa*	*Melissodes*
Eufriesea	–	0.12362	0.09264	0.12338	0.33947	0.31960	0.32128	0.28283	0.28327	0.28032	0.26612	0.44971	0.41273	0.42038	0.37622	0.42674
Euglossa	0.05337	–	0.11631	0.14354	0.30952	0.31417	0.28571	0.29167	0.30382	0.29314	0.27564	0.46798	0.39738	0.41071	0.37500	0.43352
Eulaema	0.03486	0.05110	–	0.13536	0.33749	0.33898	0.31299	0.29984	0.29435	0.28503	0.27754	0.44070	0.39289	0.39936	0.38057	0.44287
Exaerete	0.05548	0.07009	0.05546	–	0.35918	0.35887	0.32914	0.29946	0.31832	0.30726	0.28920	0.45031	0.39677	0.41281	0.37068	0.43668
Bombus avi.	0.14375	0.13963	0.13942	0.15658	–	0.09503	0.07738	0.27381	0.24228	0.28590	0.25180	0.37043	0.38601	0.37500	0.36310	0.39424
Bombus pen.	0.12544	0.12998	0.12771	0.14498	0.04265	–	0.05080	0.28698	0.26823	0.27970	0.24997	0.34880	0.35421	0.31910	0.36297	0.36783
Bombus ter.	0.12757	0.12569	0.12111	0.13661	0.03593	0.01930	–	0.26190	0.24215	0.26074	0.22795	0.31777	0.32529	0.31548	0.35119	0.35677
Lestrimel.	0.13100	0.13944	0.13296	0.14231	0.11382	0.11714	0.10783	–	0.06072	0.06946	0.06597	0.38869	0.36776	0.36310	0.34524	0.36557
Scaptotrig.	0.13707	0.14400	0.13699	0.15280	0.10958	0.11687	0.10759	0.03041	–	0.05678	0.04947	0.37520	0.35750	0.35766	0.33964	0.40030
Tetragona	0.12906	0.13765	0.12673	0.14245	0.12098	0.11310	0.11057	0.03380	0.03144	–	0.04470	0.37948	0.34588	0.36982	0.36187	0.39903
Trigona	0.12914	0.13785	0.12913	0.14266	0.11218	0.11070	0.10220	0.03600	0.03256	0.02331	–	0.36935	0.33507	0.35940	0.34734	0.39206
Apis mel.	0.19213	0.19828	0.18516	0.19483	0.17263	0.15866	0.15071	0.18233	0.18220	0.17770	0.18021	–	0.10672	0.09886	0.37535	0.40039
Apis nig.	0.18318	0.18038	0.17182	0.18157	0.18436	0.16359	0.15683	0.18228	0.18241	0.17217	0.17614	0.03699	–	0.04380	0.37157	0.45374
Apis dor.	0.18663	0.18725	0.17601	0.18826	0.16751	0.14284	0.14359	0.16733	0.17009	0.16995	0.17176	0.04332	0.03003	–	0.38095	0.40267
Xylocopa	0.15375	0.15936	0.15143	0.15838	0.16174	0.15345	0.15377	0.15139	0.15634	0.15621	0.15580	0.16530	0.16694	0.17131	–	0.35248
Melissodes	0.18157	0.18296	0.18305	0.18571	0.16804	0.15482	0.14928	0.16753	0.17211	0.17266	0.17205	0.16548	0.18740	0.17077	0.15132	–

respectively, for 1st, 2nd, and 3rd positions, indicating no significant heterogeneity of frequencies among the taxa examined. Table 4 shows the uncorrected pairwise divergences between sequences, for all nucleotides and for 3rd codon position sites alone. As expected, most of the sequence variability (85% of phylogenetically informative changes) occurs at 3rd positions (Fig. 1), where changes are usually synonymous. From the TI/TV plot in Fig. 2, a TI/TV ratio of 4.3 was estimated for 3rd positions from a regression slope calculated from the far left portion of the graph (TV ≤ 8), where TI are unlikely to be saturated. After approximately 8 TV, the number of TI in 3rd positions appears to remain constant relative to TV, which suggests saturation of TI at those sites.

Test of Orthology of the Corbiculate Bee LW Opsins

Simultaneous analysis of the corbiculate bee opsin sequences with other insect opsin sequences retrieved from GenBank resulted in a single most-parsimonious tree (Fig. 3). Note that all the corbiculate bee opsins sequenced in our study form a monophyletic group (BV 94%) distinct from the other insect opsins. Importantly, our *A. mellifera* sequence appears as sister group to the *A. mellifera* sequence described as LW *Rh* (green) by Chang *et al.* (1996), identifying this gene copy as the one we have sequenced. The branch subtending these two *Apis* sequences (Fig. 3) has a BV of 99% rather than 100% because the Chang sequence, representing the entire gene, comprised approximately twice the number of nucleotides. All groupings in this topology are entirely concordant with results from Townson *et al.* (1998), including the clustering of bee LW *Rh* with ant (1 and 2) opsin 1, and the mantid and the moth opsin, which in turn is a sister clade (100% BV) to the paralogous *Drosophila Rh1* and *Rh2* opsins. A similar

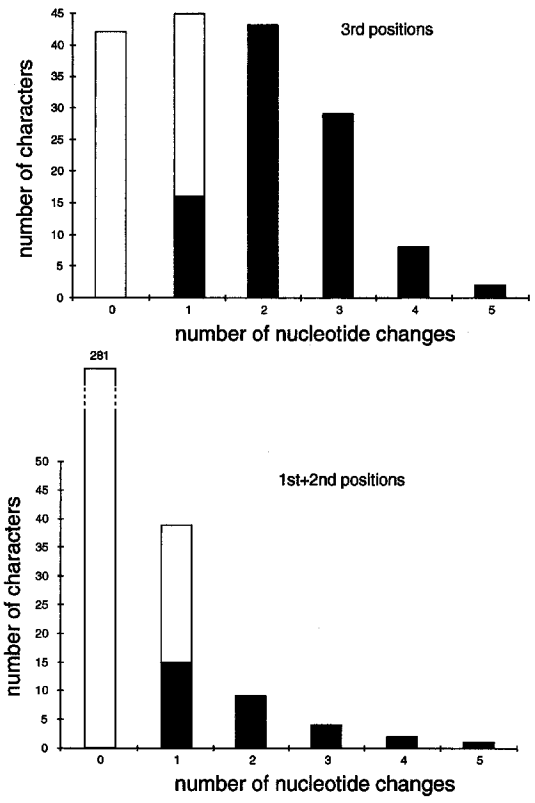

FIG. 1. Frequency distribution of the number of nucleotide changes per character, identified by 3rd and 1st + 2nd codon positions. Shaded bars represent parsimony-informative characters.

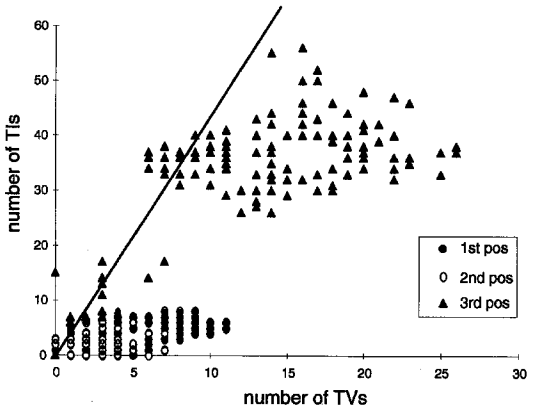

FIG. 2. Number of transitions (TI) plotted against the number of transversions (TV) for all pairwise comparisons for 1st, 2nd, and 3rd codon positions. Note that the number of TI + TV = total pairwise distance. The regression slope was calculated for 3rd-position pairwise divergences for which the number of TV ≤8.

pattern of clustering of LW insect opsins (minus the bees) was reported by Crandall and Cronin (1997).

Corbiculate Bee Tribal Phylogeny

The strict consensus of the nine MP trees resulting from an unweighted parsimony analysis of the nucleotide sequence data is shown in Fig. 4A. Because saturation is suggested at 3rd positions (Fig. 2) and a TI/TV ratio of 4.3 was inferred, we performed a weighted parsimony analysis, assigning TI at 3rd positions a weight of 1 and all other substitutions a weight of 4. The strict consensus of the six resulting MP trees is shown in Fig. 4B.

In other weighting schemes, we excluded 3rd or 1st + 2nd positions entirely. Exclusion of 3rd positions (using only 1st + 2nd) in an unweighted parsimony analysis resulted in 64 trees; the strict consensus tree (figure not shown) is fully concordant with the tree in Fig. 4A. Exclusion of 1st + 2nd positions (using only 3rd positions with TV = 4 TI) resulted in a topology that was also fully concordant with (except for the placement of one taxon inside the Bombini) and more resolved (relationships among Meliponini resolved) than the tree in Fig. 4A. The ILD test resulted in a P value of 0.36, indicating no significant incongruence between these two classes of sites despite the different constraints operating on each. Although most of the nucleotide variation occurs at 3rd positions (Fig. 1), 1st and 2nd positions nevertheless contain substantial phylogenetic information (i.e., they reconstruct the same tribal clades, (Meliponini + Bombini) and (Meliponini + Bombini + Euglossini), as those inferred from analysis of all positions).

The ML analysis, which provides another method of accounting for saturation of substitutions, resulted in the tree shown in Fig. 5. The tribal topology is the same as that estimated from parsimony (Figs. 4A, 4B), except that the outgroup taxon *Melissodes* falls within Apini (discussed below).

These results all strongly support the monophyly of each of the four corbiculate apine tribes. The trees estimated from the weighted MP and the ML analyses show a more strongly supported relationship between Bombini + Meliponini (87% and 71% BV, respectively) than those estimated from unweighted parsimony (46% BV). The low support for this clade on the unweighted MP tree is likely to be the result of saturation of transitions at 3rd positions. However, on neither the weighted MP tree nor the ML tree is the monophyly of the corbiculate apines strongly supported. We have investigated this problem using Relative Apparent Synapomorphy Analysis (Lyons-Weiler *et al.*, 1996), which allows the detection of long branches in sequence data (Lyons-Weiler and Hoelzer, 1997). The results of the RASA analyses identified the presence of two long branches in the estimated phylogeny, one leading to the tribe Apini and one leading to the outgroup *Melissodes*. Because long branches have a tendency to attract one another and thereby mislead tree estimation, the clustering of *Apis* and *Melissodes* on the ML tree might be explained by the attraction of the two detected long branches. This rooting problem is not restricted to the LW *Rh* opsin gene. We have encountered it with other

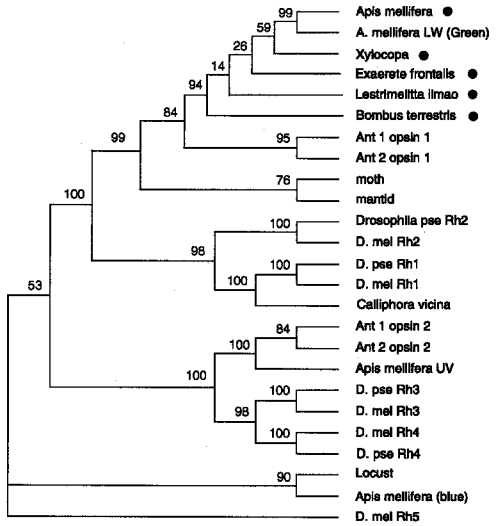

FIG. 3. Phylogenetic relationships between different insect opsin sequences, including sequences obtained in this study from each tribe of the corbiculate bees (represented by closed circle). The tree was estimated using unweighted parsimony analysis of amino acid sequences, implemented in PAUP*4 (one most-parsimonious tree). Numbers on the tree represent bootstrap values based on 400 replications.

genes used for estimating relationships among the corbiculate bees and will discuss it in detail elsewhere (Cameron and Mardulyn, unpublished data).

DISCUSSION

Phylogenetic analyses of opsins have been limited to examining relationships among different gene copies of the opsin multigene family (Carulli and Hartl, 1992; Carulli *et al.*, 1994; Yokoyama, 1995; Chang *et al.*, 1995, 1996; Townson *et al.*, 1998) and to exploring diversification and convergence in gene function (Yokoyama and Yokoyama, 1990, 1996; Crandall and Cronin, 1997). The results of our novel use of LW *Rh* to estimate phylogenetic relationships among bees provide a convincing example that at least one of the insect visual pigment genes can be used for reconstructing organismal phylogenies. Additional opsin genes are present in *A. mellifera* (Townson *et al.*, 1998), including blue- and UV-sensitive opsins, which could prove potentially useful for phylogenetic reconstruction of taxa at the level of tribes and subfamilies or other taxonomic levels. These genes require examination for such purposes.

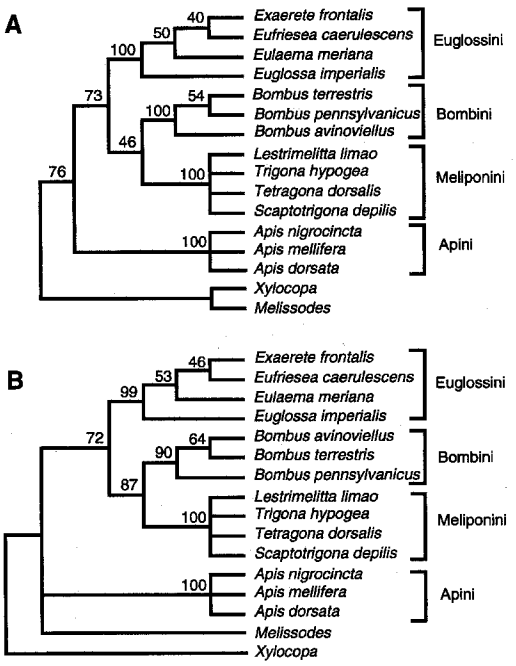

FIG. 4. Phylogenetic trees of the four tribes of corbiculate bees estimated using unweighted (A) and weighted (B) parsimony analysis (branch-and-bound search) of all nucleotides. In the weighted analysis, TI in 3rd positions were assigned a weight of 1; all other substitutions were assigned a weight of 4. Topology (A) is the strict consensus of nine equally parsimonious trees (total number of characters = 502; parsimony-informative characters = 132; Tree Length = 363; CI = 0.658; RI = 0.742). Topology (B) represents the strict consensus of six equally parsimonious trees (total number of characters = 502; parsimony-informative characters = 130; Tree Length = 832; CI = 0.706; RI = 0.782). Numbers above internal nodes represent bootstrap values (400 replications).

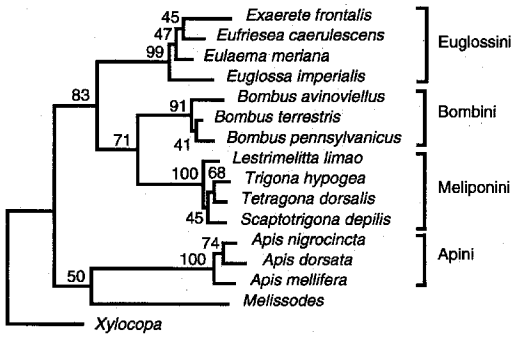

FIG. 5. Phylogenetic tree of the corbiculate tribes estimated using the maximum likelihood method (100 random addition sequences and TBR swapping) implemented with the HKY85 + Γ_4 model of DNA substitution (Hasegawa *et al.*, 1985; Yang, 1993). The maximum likelihood estimates of the TI:TV ratio, shape parameter of the gamma distribution, and proportion of invariable sites are 2.62, 0.86, and 0.34, respectively. Numbers beside internal nodes are bootstrap values.

One critical test of the accuracy of a phylogenetic hypothesis is to demonstrate the taxonomic congruence of trees estimated from multiple independent lines of evidence (Hillis, 1995; Miyamoto and Fitch, 1995; Friedlander *et al.*, 1996; Mitchell *et al.*, 1997; but see DeSalle and Brower, 1997). Notably, the corbiculate bee MP tree estimated from the LW *Rh* nucleotide sequences is entirely congruent (same branching pattern) with other tribal reconstructions of the corbiculate bees estimated from other genes, both mitochondrial 16S (Cameron, 1993) and cytochrome *b* (Koulianos, 1998) and the nuclear 28S gene (Sheppard and McPheron, 1991). The taxonomic congruence that we find with the opsin sequences suggests that LW *Rh* has retained genealogical information useful for reconstructing phylogenetic relationships among apine tribes.

If a gene is highly conserved at 1st + 2nd codon positions but changes rapidly (exhibiting saturation of TI) at 3rd positions, it may not be useful for phylogeny reconstruction at intermediate or higher taxonomic levels. The LW *Rh* gene contains information useful at both 3rd and 1st + 2nd position sites, at least for tribes and subfamilies of bees. Excluding 3rd positions, which were six times more variable than 1st + 2nd positions combined and which contained most of the phylogenetically informative sites, resulted in a partially resolved tree that was concordant with the tree estimated using all of the data or only 3rd-position sites. Likewise,

accounting for saturation by downweighting TI in 3rd positions improved node support in general, but had little influence on the tree topology (Fig. 4B). Thus, the tree is stable to different models of evolution and to various weighting procedures, and useful information can be obtained from the LW *Rh* gene at several levels of site variability.

In addition to showing the phylogenetic usefulness of the LW *Rh* gene, this work adds a new gene to the accumulating molecular evidence suggesting a sister group relationship between Meliponini and Bombini, and a clade (Meliponini + Bombini + Euglossini). These relationships are important for understanding the evolution of social behavior in bees because they suggest that the two advanced eusocial groups, Apini and Meliponini, evolved independently along different lineages. It is interesting, if not provocative, that all of the molecular data sets provide results that contradict the phylogeny estimated from morphological data (Roig-Alsina and Michener, 1993). In contrast to the dual origin inferred from molecular data, a single origin is inferred from the morphology, such that the advanced eusocial Apini and Meliponini are sister taxa within the clade ((Meliponini + Apini) + Bombini). These conflicting results are of concern and have motivated a comprehensive investigation of congruence of all available molecular and morphological data for the corbiculate bees (Cameron and Mardulyn, in progress).

ACKNOWLEDGMENTS

We thank Rob DeSalle and Donat Agosti for the LWRh-F and LWRh-R primer sequences and Dave Swofford for use of test versions of PAUP*4, and we are indebted to Stella Koulianos, Gard Otis, Regula Schmid-Hempel, Horst Schwarz, and David Roubik for assistance with taxon collection. We especially thank Belinda Chang and Keith Crandall for helpful discussions and Belinda for sharing her *A. mellifera* blue and UV opsin sequences. We gratefully acknowledge Jim Whitfield, Rob DeSalle, Andy Brower, and an anonymous reviewer for their careful reading of an earlier version of the manuscript. This research was supported by NSF (GER-9450117) and USDA (NRICGP-9501893) Grants to S.A.C. and J. Whitfield, respectively.

REFERENCES

Alexander, B. A. (1991). Phylogenetic analysis of the genus *Apis* (Hymenoptera: Apidae). *Ann. Entomol. Soc. Am.* **84:** 137–149.

Brower, A. V. Z., and DeSalle, R. (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.

Camargo, J. M. F., and Pedro, S. R. (1992). Systematics, phylogeny and biogeography of the Meliponinae (Hymenoptera, Apidae): A mini-review. *Apidologie* **23:** 509–102.

Cameron, S. A. (1993). Multiple origins of advanced eusociality in bees inferred from mitochondrial DNA sequences. *Proc. Natl. Acad. Sci. USA* **90:** 8687–8691.

Cameron, S. A., Derr, J. N., Austin, A. D., Wooley, J. B., and Wharton, R. A. (1992). The application of nucleotide sequence data to phylogeny of the Hymenoptera: A review. *J. Hymenop. Res.* **1:** 63–79.

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

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

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

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

Chase, M. R., Bennett, R. R., and White, R. H. (1996). Expression of opsin mRNA in normal and vitamin A deficient retinas of the sphingid moth *Manduca sexta*. *Vis. Neurosci.* **13:** 353–358.

Cho, S., Mitchell, A., Regier, J. C., Mitter, C., Poole, R. W., Friedlander, T. P., and Zhao, S. (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., Hall, K. J., Wilson, D. B., Wideman, C. V., Townson, S. M., Chadwell, L. V., and Britt, S. G. (1996). Identification of a novel *Drosophila* opsin reveals specific patterning of the R7 and R8 photoreceptor cells. *Neuron* **17:** 1101–1115.

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

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

Danforth, B. N., and Ji, S. (1998). Elongation factor-1α occurs as two copies in bees: Implications for phylogenetic analysis of EF-1α sequences in insects. *Mol. Biol. Evol.* **15:** 225–235.

DeSalle, R., and Brower, A. V. Z. (1997). Process partitions, congruence, and the independence of characters: Inferring relationships among closely related Hawaiian *Drosophila* from multiple gene regions. *Syst. Biol.* **46:** 751–764.

Fang, Q. Q., Cho, S., Regier, J. C., Mitter, C., Matthews, M., Poole, R. W., Friedlander, T. P., and Zhao, S. (1997). A new nuclear gene for insect phylogenetics: Dopa decarboxylase is informative of relationships within Heliothinae (Lepidoptera: Noctuidae). *Syst. Biol.* **46:** 269–283.

Farris, J. S., Källersjö, M., Kluge, A. G., and Bult, C. (1994). Testing significance of incongruence. *Cladistics* **10:** 315–319.

Felsenstein, J. (1978). Cases in which parsimony and 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.

Friedlander, T. P., Regier, J. C., Mitter, C., and Wagner, D. L. (1996). A nuclear gene for higher level phylogenetics: Phosphoenolpyruvate carboxykinase tracks mesozoic-age divergences within Lepidoptera (Insecta). *Mol. Biol. Evol.* **13:** 594–604.

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

Gupta, R. S. (1995). Phylogenetic analysis of the 90kD heat shock family of protein sequences and an examination of the relationships among animals, plants, and fungi species. *Mol. Biol. Evol.* **12:** 1063–1073.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22:** 160–174.

Hendy, M. D., and Penny, D. (1989). A framework for the quantitative study of evolutionary trees. *Syst. Zool.* **38:** 297–309.

Hillis, D. M. (1995). Approaches for assessing phylogenetic accuracy. *Syst. Biol.* **44:** 3–16.

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

Huelsenbeck, J. P. (1997). Is the Felsenstein zone a fly trap? *Syst. Biol.* **46:** 69–74.

Kim, J. (1996). General inconsistency conditions for maximum parsimony: Effects of branch lengths and increasing numbers of taxa. *Syst. Biol.* **45:** 363–374.

Kimsey, L. S. (1987). Generic relationships within the Euglossini (Hymenoptera: Apidae). *Syst. Entomol.* **12:** 63–72.

Koulianos, S., Schmid-Hempel, R., Roubik, D. W., and Schmid-Hempel, P. (1999). Phylogenetic relationships within the corbiculate Apinae (Hymenoptera) and the evolution of ensociality. *J. Evol. Biol.* **12:** 380–384.

Lyons-Weiler, J. (1998). RASA 2.2 and Manual for the Macintosh. [http://loco.biology.unr.edu/archives/rasa/rasa.html].

Lyons-Weiler, J., Hoelzer, G. A., and Tausch, R. J. (1996). Relative Apparent Synapomorphy Analysis (RASA). I. The statistical measurement of phylogenetic signal. *Mol. Biol. Evol.* **13:** 749–757.

Lyons-Weiler, J., and Hoelzer, G. A. (1997). Escaping from the Felsenstein zone by detecting long branches in phylogenetic data. *Mol. Phylogenet. Evol.* **8:** 375–384.

Maddison, W. P. and Maddison, D. R. (1992). MacClade: Analysis of Phylogeny and Character Evolution. Sinauer, Sunderland, MA.

Michener, C. D. (1974). "The Social Behavior of the Bees," Harvard Univ. Press, Cambridge, MA.

Michener, C. D. (1990). Classification of the Apidae (Hymenoptera). *Univ. Kansas Sci. Bull.* **54:** 75–164.

Milinkovitch, M. C., LeDuc, R. G., Adachi, J., Farnir, F., Georges, M., and Hasegawa, M. (1996). Effects of character weighting and species sampling on phylogeny reconstruction: A case study based on DNA sequence data in cetaceans. *Genetics* **144:** 1817–1833.

Mitchell, A., Cho, S., Regier, J. C., Mitter, C., Poole, R. W., and Matthews, M. (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., and Fitch, W. M. (1995). Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* **44:** 64–76.

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

O'Tousa, J. E., Baehr, W. B., Martin, R. L., Hirsh, J., Pak, W. L., and Applebury, M. L. (1985). The *Drosophila* ninaE gene encodes an opsin. *Cell* **40:** 839–850.

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

Roig-Alsina, A., and Michener, C. D. (1993). Studies of the phylogeny and classification of long-tongued bees (Hymenoptera: Apoidea). *Univ. Kansas Sci. Bull.* **55:** 124–162.

Sheppard, W. S., and McPheron, B. A. (1991). Ribosomal DNA diversity in Apidae. *In* "Diversity in the Genus *Apis*" (D. R. Smith, Ed.), pp. 89–102. Westview, Boulder, CO.

Swofford, D. L., Olsen, G. J., Waddell, P. J., and Hillis, D. M. (1996). Phylogeny reconstruction. *In* "Molecular Systematics" (D. M. Hillis, C. Moritz, and B. K. Mable, Eds.), 2nd ed., pp. 407–514. Sinauer, Sunderland, MA.

Swofford, D. L. *PAUP** 4.0. 1998. [(test versions d63-d64). Washington, DC, Laboratory of Molecular Systematics, Smithsonian Institution.] Sinauer Associates, Sunderland, MA, [in press].

Thomas, R. H., and Hunt, J. A. (1993). Phylogenetic relationships in *Drosophila:* A conflict between molecular and morphological data. *Mol. Biol. Evol.* **10:** 362–374.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

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

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

Waters, E. R. (1995). An evaluation of the usefulness of the small heat shock genes for phylogenetic analysis in plants. *Ann. Mo. Bot. Gard.* **82:** 278–295.

Williams, P. H. (1998). An annotated checklist of bumble bees with an analysis of patterns of description (Hymenoptera: Apidae, Bombini). *Bull. Nat. Hist. Mus. Lond. (Entomol.)* **67:** 79–152.

Yang, Z. (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* **10:** 1396–1401.

Yokoyama, S. (1995). Amino acid replacements and wavelength absorption of visual pigments in vertebrates. *Mol. Biol. Evol.* **12:** 53–61.

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

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

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